Law Faculty Scholarship

1992

Lawyer Decision Making: The Problem of Prediction

Marjorie McDiarmid

Follow this and additional works at: https://researchrepository.wvu.edu/law_faculty

Part of the Legal Profession Commons
This Article examines three competing models for lawyer decision making. Reviewing literature drawn from other disciplines, Professor McDiarmid applies each model to a particular lawyer decision task and provides a critique both of applicability and of the underlying assumptions of the models themselves. The Article concentrates on the problem of prediction in the face of uncertainty.

I. INTRODUCTION

As they portrayed it on *American Playhouse,*

1. the exchange took place between the lawyer and his clients in the courtroom holding cell. An inspired script writer has one of the clients, I think it was Loeb, say “if [you’re wrong], we are meat on a hook.” And they literally would have been, for those were hanging times.

The lawyer was Clarence Darrow and the clients were Leopold and Loeb, at that point accused of the heinous murder of Bobby Franks. The decision that Darrow was advocating was the entry of a plea of guilty, thereby waiving trial by jury and staking his clients’ lives on Darrow’s presentation to Judge Caverly alone. It turns out to have been the right decision, of course. The summation is too deterministic for my

---

* Professor of Law, West Virginia University. B.A., University of Rochester; J.D, Columbia University; LL.M., Harvard University. I wish to acknowledge the support and editorial assistance of Professors Thomas Barton, Gary Bellow, Donald Gifford, and Jeffrey Lewin. Any mistakes are mine.


2. Nathan Leopold, Jr. and Richard Loeb were the sons of millionaires in Chicago. Leopold was 19 and Loeb was 18 at the time of the trial in 1924. Bobby Franks, a cousin of Loeb’s, was 14 years old at the time they killed him. There was strong physical evidence linking the young men to the crime and both confessed. They referred to the killing as taking place for the “sake of a thrill.” Both defendants had planned to enter law school in the fall. Leopold had already been admitted to Harvard.

For discussions of the case and of Darrow’s decision, see IRVING STONE, CLARENCE DARROW FOR THE DEFENSE 241-73 (1958) and ATTORNEY FOR THE DAMNED 16-88 (Arthur Weinberg ed., 1989) [hereinafter DAMNED]. DAMNED contains the edited transcript of Darrow’s actual summation.
taste and the psychology too Freudian, but Darrow knew his case, his judge and the temper of his times. The young men did not die. Most lawyer decisions, to be sure, do not carry the kind of consequences Darrow’s did. But if the best empirical information about legal practice is even approximately correct, all transactional legal matters and the vast majority of legal disputes are settled not through the intervention of any state apparatus, but rather through the decision making and concomitant advice-giving activity of lawyers. The lawyer’s discharge of the counseling function is virtually unstudied, yet it forms the recognized core of much lawyer activity. It is for good reason that “counselor” or “counsel” is the most widely used honorific for lawyers in formal address.

A word on the purposes of this essay. They are two: speculative and normative. There are many facets of lawyer decision making about which little is known. It would be useful to know whether the decision making power allocated by professional norms variously to the lawyer and the client in fact rest where they are assigned. It would be interesting

3. See DAMNED, supra note 2.

4. Judge John R. Caverly heard Darrow’s three day summation and sentenced the defendants to life imprisonment plus 99 years. His announced ground for the decision was that both defendants were minors. Loeb was stabbed to death in prison in 1936. Leopold was paroled in 1958 and lived out his life in Puerto Rico. While in prison he was a subject in medical experiments involving malaria. His work after his release has been variously described as scientific research in the field of communicable disease, see Richard Cohen, Clarence Darrow’s Lesson, WASH. POST, Dec. 2, 1988, at A27, and as that of a hospital technician for $10 per week, see File in 1924 “Thrill Slaying” Case Found, UNITED PRESS INT’L, Oct. 10, 1985. He died August 29, 1971 of natural causes.

5. Contract and will drafting, partnership and corporation creation, conveyancing and the like.

6. The best quantitative data on civil law practice in the United States were collected by the Civil Litigation Research Project funded by the Office for Improvements in the Administration of Justice of the Department of Justice. Initial results were reported in 1981. That study showed that less than half of all disputes brought to lawyers ended in a case being filed. Richard E. Miller & Austin Sarat, Grievances, Claims, and Disputes: Assessing the Adversary Culture, 15 L. & Soc’y Rev., 525, 537, 542-43 (1981). Further, data from the Federal Judicial Workload Statistics Profile for the years from 1985 to 1990 show that on average 473 cases per judge, both civil and criminal, are filed annually while only an average 35 per judge reach trial on an annual basis. Thus, as a rough measure, close to 13% of federal cases filed reach trial. The others are either disposed of by motion (court action), abandoned, or settled (lawyer/client decision). Unfortunately, there are no data breaking out these non-trial dispositions. See STATISTICAL ANALYSIS & REPORTS DIV. ADMINISTRATIVE OFFICE OF THE UNITED STATES COURTS, FEDERAL JUDICIAL WORKLOAD STATISTICS.

7. For one of the only empirical studies on the allocation of lawyer and client decision authority, see Douglas E. Rosenthal, LAWYER AND CLIENT: WHO’S IN CHARGE? (1977).
to find out whether lawyers and clients experience the same problems that plague other decision makers in valuing outcomes. It would be particularly helpful to conduct an empirical study of how lawyers actually arrive at decisions. The exploration of ideas conducted here is a prerequisite to the resolution of these issues.

This work represents a theoretical analysis of decision making in which the practitioner must select action in the face of uncertainty. Darrow did not know how the judge would decide his clients’ fate. He could make at best an educated guess. Within the domain of decision making under uncertainty, I opt to further confine my study to methods that decision makers have available to structure and if possible to reduce that uncertainty, i.e., methods of prediction.

I have isolated three prediction methodologies which should be of assistance to lawyers: decision theory, inductive probability, and causal decision making. These methods, drawn from the realms of economics, psychology, scientific investigation, and philosophy, might capture how lawyers in fact decide. More likely, the methods idealize aspects of decision making that are used much more informally in real life decision processes. Or it is possible that they might be completely foreign to present methods of lawyer decision making. Because of the lack of empirical data, the descriptions of lawyer decision making offered here are simply a set of plausible hypotheses from which to begin the research required. As I will describe in Section IV.B, without such a set of hypotheses, empirical research is rudderless.

But there is additional virtue in this discussion now, even if the descriptive value of the material is speculative. The empirical study I propose necessarily will be a protracted and uncertain process. Lawyers need to make predictions in the meantime. Thus, the normative value of this Article lies in acquainting those lawyers with the hard-won insights of other thinkers on decision making so that lawyer prediction may be

---

8. The fact that not all decision makers attach the same value to even as consistent a commodity as money has long been recognized. As long ago as the 1700s, Daniel Bernolli developed a theory of utility that equated values assigned by different decision makers by weighting monetary sums by the worth of those sums to the decision maker involved. For a general discussion of utility concepts, see Howard Raiffa, Decision Analysis: Introductory Lectures on Choices Under Uncertainty (1968). More recently, it has become clear that decision makers do not exhibit consistent utility curves, with the result that values vary not only as between decision makers but also across situations faced by the same decision maker. The primary researchers in this area have been Amos Tversky and Daniel Kahneman. For an example of their work, see Rational Choice and the Framing of Decisions, in Decision Making: Descriptive, Normative and Prescriptive Interactions 167-92 (David E. Bell et al. eds., 1988) [hereinafter Decision Making]. For a summary of developments in the field of utility generally, see J. Frank Yates, Judgment and Decision Making (1990).
improved right now. If lawyers in fact use prediction methods similar to those described here, the profession will benefit from reflections concerning their advantages and pitfalls. If lawyers do not currently use these methods, perhaps they should. At least the choice not to use them should be informed.

An incidental benefit that is not quite so utilitarian lies in the specification of the models themselves. Drawn from a wide variety of sources, they have never, to my knowledge, been contrasted as they are here.

Why three predictive methods? Most simply because predictive tasks differ. I would characterize them alliteratively as follows. Decision theory has as its domain the evaluation of relatively well-defined problems. Inductive probability extrapolates from known data to forecast less well-known outcomes. Causal decision making explains the future in light of previously identified cause-and-effect relations.

The evaluative function served by decision theory presupposes a clear set of options coupled with a need to weigh the outcomes associated with each. Although some decision theorists expand the use of the method to arenas in which hard prior probability data are lacking, the method calls for the assignment of precise values to likely outcomes. Decision theory is also mathematically rigorous in its manipulation of these values. It uses a probability calculus composed of mathematical operations to transform assigned probabilities into outcome predictions. The most common use of decision theory methodology by lawyers is in the weighing of the relative merits of settlement versus trial. I will use an example of such a problem to illustrate my discussion below. There the decision maker is evaluating known options: settlement, motion to dismiss, verdict. That lawyer can use past experience to assign numerical likelihood predictions to these outcomes. The probability calculus will combine these judgments and provide an assessment of the most favorable course of action. Decision theory accommodates most comfortably well defined problems posing clear alternative solutions.

By contrast, there would seem to be large classes of legal decision problems in which options are not clearly apparent. Further, if those options do emerge in the course of problem solving, it is not evident what likelihood of success attends them. An example which I use below of such a less clearly defined problem is the dilemma posed to a lawyer deciding whether to take on a case in any uncertain area of the law. Too many possible outcomes—inability to find qualified witnesses, unfavorable reading of ambiguous statutory terms, ill-timed legislative activity, failure to secure necessary material in discovery, etc.—complicate this problem to make a decision theoretic approach useful. The case might fail from myriad combinations of factual circumstances, legal rulings, procedural vicissitudes, and political judgments. No good numbers attach to these
combinations because no case in precisely these circumstances has gone before.

Rather than pursue the illusory rigor of decision theory in this instance, inductive probability, the second approach considered here, would cause the lawyer to attempt to define broadly the variables which should affect her judgment and then give each of these variables a gatekeeper function. If the novel cause of action required a showing of intentional misconduct on the part of the alleged tortfeasor, then evidence of the requisite intent would be a relevant variable, and the absence of such intent would dictate that the lawyer cease her analysis with the conclusion that the cause of action was untenable. This use of the pass-fail, gate-keeper approach, of course, makes inductive probability a blunter instrument than is decision theory, which recognizes gradations of probability. But where precise data on a new or poorly defined problem are lacking, it permits a more realistic assessment mechanism. It permits the decision maker to extrapolate from available knowledge to a useful, though gross, judgment about problematic outcomes.

The first two predictive methods, decision theory and inductive probability, are thus distinguishable by the precision of both the problems they address and the data they use. But both seek to produce a similar judgment about the future: how likely it is that a particular outcome will occur. In reaching this judgment, most practitioners of both methods use the raw data of prior experience as their guide. What has happened in the past is deemed likely to recur simply by dint of that prior experience.

The final method of decision making also builds on prior correlations between choices and outcomes, but it does not stop there. Rather than saying that the judge is likely to decide a cause in a certain manner because she has done so with some frequency in the past, causal decision theory looks beyond mere correlation and attempts to develop explanation. The judge decides this kind of case a certain way because she is predisposed by her education and values to find certain arguments compelling. The obvious advantage of causal thinking is that, if the causal attribution is correct, it will enable the decision maker to predict accurately even a low probability event. The judge will decide this particular case differently from the way she decides most such cases because of the presence or absence of the critical causal factor. Yet, as I will describe, there are risks in causal thinking undertaken in too facile a manner which argue that a pure probability approach may in many circumstances be more accurate.

At any rate, I would contend that the lawyer community stands to benefit from the differentiation and study of the modes, benefits and drawbacks of all three methods of decision making. In my conclusion to this discussion, I will revisit the question of which, if any, of these methods are likely to be used by and merit the use of that community.
II. DECISION THEORY

A. Introduction

One approach to decision making is so dominant that its name seeks to pre-empt the field. Referred to uniformly as "decision theory" or "decision analysis," this methodology calls its practitioners to assign probabilities and utilities\(^9\) to events. It then prescribes the combination of these assignments to determine which course of action is to be preferred. Because this essay deals solely with the predictive portion of decision making, I will discuss probability here and leave utility issues for another day. I will describe first the "decision theory" methodology, then set forth some of its less obvious constraints.

B. Methodology

Let me posit for purposes of this discussion a settlement decision in which the lawyer wishes to advise her client. This is a personal injury action arising from an automobile collision. The injured plaintiff has fully recovered from those injuries after a prolonged course of hospital treatment, recuperation, and physical therapy. That course of treatment consumed two years and cost $200,000 in actual expenditures. In addition, the plaintiff has lost wages of $150,000 for the same period. The plaintiff and his wife are seeking $1,050,000 for pain and suffering and loss of consortium. Assume further that the plaintiff sustained property damage of $10,000. Thus, the total damage claim is $1,410,000. The defendant's carrier, on the eve of trial, has offered to settle the case for $800,000. The plaintiff and his wife seek the advice of the lawyer on whether they should accept the offer.

It is the lawyer's judgment that her case will almost certainly survive a motion for a directed verdict or a motion for a judgment of law at all stages during the trial. It is further her belief that a jury is likely to find

---


10. Utilities are the values that the decision maker assigns to various outcomes. See supra note 8. Since not all people value even money uniformly, some method of quantifying value must be employed. Because this Article avoids exploration of utilities, however, I will employ throughout the fiction that the decision maker assigns face value to monetary outcomes. A decision maker employing this utility structure is said to use "expected monetary value." See Raiffa, supra note 8, at 8-9.
liability on the part of the defendant, although she is less certain of this than of the judgment that the case will reach the jury. Finally, if the case goes to the jury and if they find the defendant liable, she rates the most likely monetary award as $1,000,000, although she would not be badly surprised by a verdict of $1,410,000 or one for $750,000. She thinks that it is very unlikely that the jury will come back, assuming a finding of liability, with an amount of $500,000. In her view, any other verdict is impossible. 11 How, in line with decision theory, should the lawyer determine the meaning of her judgments so that she can advise her clients of the probable outcome of their case? 12

The first chore that confronts the lawyer under decision theory is the conversion of her likelihood judgments into numeric probabilities. 13 There are several constraints on these assignments which I will merely note now. The assignments must fall between zero and one, with zero representing no chance that the event will occur and one representing certainty.

\begin{equation}
0 \leq P(X) \leq 1
\end{equation}

At each decision point, all of the probabilities assigned to possible outcomes must sum to one.

\begin{equation}
P(X_1) + P(X_2) + P(X_3) + \ldots + P(X_n) = 1
\end{equation}

11. See infra Section II.D to see why this assumption is made.
12. I will not address here the method by which this information should be conveyed. I will address simply how the figures should be computed.

The argument that there is too much variability in non-numeric expressions of probability is supported. See Robin M. Hogarth, Judgement and Choice: The Psychology of Decision 190-93 (2d ed. 1987). Yates, supra note 8, at 30, points to studies, for example, in which doctors assigned numeric equivalents, ranging from 35% to 77%, to the clause "significant chance." There is a great deal of slippage in non-numeric probability statements. But see Michael Smithson, Ignorance and Uncertainty: Emerging Paradigms 164-68 (1989) (positing that data in numeric form may not be as useable as experiments have assumed). See also T.J. Nagy & L.J. Hoffman, Exploratory Evaluation of the Accuracy of Linguistic versus Numeric Risk Assessment of Computer Security, Technical Report GWU-11ST-81-07 (George Washington University Computer Security Research Group, 1981) (finding that analysts using linguistic assessments performed better largely because they avoided the extremes of inaccurate judgments), cited in Smithson, supra, at 149.
Thus, the choices given at each decision point must be exhaustive and exclusive. There cannot be an unknown amount of space left for unknown contingencies. Finally, combinations of joint probabilities are accomplished by multiplying individual probabilities together.

\[
(3) \ P(X \& Y) = P(X|Y) \times P(Y)
\]

To illustrate these points with the hypothetical, at the first decision point—directed verdict/no directed verdict—the lawyer must translate "almost certainly" into some precise number between zero and one. In this case, let me suppose that she picks .95. (We will leave until much later the questions that I am sure leap to your mind: "Why .95, why not .96? Where do you get these numbers from anyway?" For now just assume our lawyer has some good basis for making the assignment that she does.) So far she has addressed the first requirement, because the number she has chosen falls between zero and one. I will call this number \( P(J_{\text{ur}}) \). Now what about the second requirement? Remember that all of the assignments at this point must total to one. So far she is not there. Her sole assignment is .95. Of course, the choice here was between two possibilities—no directed verdict/directed verdict—so if no directed verdict has a probability of .95, it is not difficult to accept that directed verdict should be assigned a probability of .05. This assignment meets our second test: .95 + .05 equaling 1. So far she is not combining decisions, so the final test is postponed. One decision down (directed verdict/no directed verdict), and two more to go: liability/no liability and amount of verdict.

It would be possible to simply jot these assignments down in any manner, but the method of choice for displaying these data among decision theorists is the decision tree. Using it will help the lawyer keep the assignments and the decision points clear and will assist later with the math. So please refer to Figure 1. From this figure you can see the situation so far. There is a settlement offer on the table of $800,000.

14. The notation used throughout this Article is as follows:
   \( P(X) = \) Probability of \( X \).
   \( P(X|Y) = \) Probability of \( X \) given that \( Y \) is true.
   \( P(X \& Y) = \) Probability of \( X \) and \( Y \).
   \( P(X) = \) Probability of not \( X \), or the contrary of \( X \).
   \( \leq \) = less than or equal to.
   \( \geq \) = greater than or equal to.

The basic decision, which is in fact the one the lawyer is attempting to evaluate, is whether to accept this offer or take the uncertain prospect of trial. Whenever the value of an option is known outright, as in the case of the settlement here, it will be shown on the tree at the end of a line. Such a value is referred to as a leaf node, thus maintaining our tree analogy (although the savvy observer will have spotted the fact that the tree has been the victim of a high wind or an indiscriminate logger and is currently reposing on its side). The line to which the leaf node is attached is called a fork. The point that connects the various forks is called a node or a branch point.

I have chosen to picture nodes in a manner different from conventional usage. Normally square nodes denote decisions and circular nodes denote chance. Since here we have two sets of decision makers—the client and the court (judge or jury)—I have used rectangles for the client decisions and diamonds for the decisions of either the judge or jury. Probabilities are entered along the fork to which they pertain. Thus, on the first decision to which the attorney has assigned probabilities, a .05 probability has been assigned to the fork for a directed verdict, which in turn has a leaf node of $0, and a probability of .95 has been assigned to the fork leading to the jury decision on liability.

The lawyer must now translate her views on the liability issue into numerical probabilities and assign them to forks with the appropriate leaf
nodes. Again there are only two options here, liability or no liability, and the two probabilities must sum to one. The lawyer chooses probabilities of \( 0.85 = P(L_{\text{liability}}) \) and \( 0.15 = P(L_{\text{no liability}}) \). These choices are added to the tree. See Figure 2.

\[
\begin{array}{c}
\text{Settle/No} \\
\text{Settle} \\
\text{Jury/No} \\
\text{Jury} \\
\text{Liability/No} \\
\text{Liability} \\
\end{array}
\]

Figure 2

Finally, the lawyer has reached the damages node. Here there are four options: $1,000,000 (most likely), $1,410,000 (not surprising), $750,000 (also not surprising) and $500,000 (very unlikely). She translates these to \( P($1_{,\text{M}}) = 0.55 \), \( P($1.41_{,\text{M}}) = 0.20 \), \( P($750_{,\text{K}}) = 0.20 \) and \( P($500_{,\text{K}}) = 0.05 \), thus obeying the law in summing to one. Figure 3 shows this situation. Again assume at this point that the lawyer has a good reason for the numbers she assigns.

I will assume for the present that every dollar is of equal value to the client in this case. Thus, dollars gained by going to trial are as valuable as dollars gained in settlement. This is unlikely to be strictly true, but as I indicated above, utility issues will have to await another article.

What the lawyer now wishes to do is compute the worth of the branches leading through trial so that she may compare that value with the

16. Assume contributory negligence is not a factor in this case. Even if it were, it would be better to depict that issue at a subsequent branch point.

17. In this Article, "M" will indicate million and "K" will indicate thousand.
Figure 3

Figure 4
$800,000 which has been offered in settlement. To do this she multiplies the value of each leaf node by the probability of the fork to which it is attached, adds together these values for each node, assigns that combined value to the previous decision node, treats that decision node as a new leaf node, and then repeats the process. Howard Raiffa, a respected teacher of decision analysis, calls this process “folding back.” Let me illustrate. Concentrate on the last set of leaf nodes at the end of the trial branch, those which set forth the likely jury verdicts in the event that liability is found. I want to determine the total value of that jury damages node. See Figure 4.

The equation looks like this:

\[
\text{Node value} = \left( P(1.4m) \times V(1.4m) \right) + \left( P(1m) \times V(1m) \right) + \left( P(750k) \times V(750k) \right) + \left( P(500k) \times V(500k) \right)
\]

Using our figures, this is \((.20 \times 1,410,000) + (.55 \times 1,000,000) + (.20 \times 750,000) + (.05 \times 500,000)\) = $1,007,000. Thus, if the case reaches the damages phase, the lawyer estimates that the value of the damage award adjusted for likelihood is slightly more than $1,000,000. But it is not certain that the case will reach that stage, so that figure must be further discounted to take trial uncertainties into account. Here the combination requirement of the third rule is illustrated. By multiplying her expected outcome by the probability of a liability finding, the lawyer is combining the probability of the verdict amount given a finding of liability by the probability of a liability finding. This is done by assigning this newly determined value to the jury damages node, treating that node as if it were now a leaf node and repeating the process again. See Figure 5.

Using the same formula, multiplying probabilities by values and adding the result, we get \((.85 \times 1,007,000) + (.15 \times 0)\) = $855,950 as the value of the liability node. The directed verdict node is then computed \((.95 \times 855,950) + (.05 \times 0)\) = $813,152.50. This is the final decision node on the trial branch. Therefore, this figure represents the fully discounted value of the jury verdict, taking into account all the uncertainties of trial. As this figure is higher than the settlement offer of $800,000, decision theory holds that the lawyer should advise her clients that they can do better by trying the case, all other things being equal.19

18. RAIFFA, supra note 8, at 21-27.
19. This, of course, assumes that going to trial or settling are free or equally costly activities. This is obviously foolish, as the cost of trial in attorneys' fees alone will almost certainly outweigh settlement costs. To account for this, the lawyer would subtract the costs of trial and settlement from their respective branches before comparing the results. This calculation, however, does not contribute to my discussion of decision
I have spent this time on the details of decision theory methodology to clarify its premises and to show its substantial value. If one accepts the premises and undertakes the valuations required, this is a highly refined and ultimately rational tool for decision making. Certainly no one should reject it on the ground that the mathematics are too difficult. Only grade school math and a clear chart are required. If, as with me, these requisites sometimes test your abilities, there are a variety of computer tools that will do at least some of the math for you.

Shifting attention now from the mechanics to the underlying theory, I will revisit the constraints that I breezed past above. The unavoidable requirements are these: 1) beliefs or judgments about likelihoods must be expressed in terms of numerical probabilities which are discrete numbers lying between zero and one; 2) those probabilities must cover the entire range of available options, i.e., they must sum to one or certainty; and 3) those probabilities are combined with the value placed on various outcomes by multiplication. Although I am using the mathematical theory, so it is omitted here.

20. Any one of the excellent computerized spreadsheets on the market will do the math flawlessly, if you plug in the proper formulas. In purely alphabetical order, see for example, Excel, Lotus 123, and Quattro. There are specialized decision tree packages as well.
structure of decision theory as the organizing scaffold for this review, I have in mind no merely mathematical quibble. As will appear, there is much philosophy buried in the numbers.

C. Precise Numeric Assignments

Is it possible to be too precise? Debate over the first constraint—assignment of numeric values—was formerly the most lively in the field of decision theory. To see why, let us examine various types of uncertainty. In the introduction, I talked about the fact that cases in which a lawyer will typically want to use some form of decision methodology will be instances in which she is trying to predict unknown future outcomes as in the settlement hypothetical. Decision theorists define at least three different types of uncertainty which are of interest: ignorance, risk, and ambiguity. Leaving ambiguity to the side for the moment, I will concentrate on ignorance and risk. Although different theorists interchange some terms, I will define these states as follows. Ignorance represents the situation in which one cannot usefully value probabilities at all. An example of an extreme case used in one text is the likelihood that the next date undertaken by a college sophomore will spark a string of consequences ending in the children fostered by him and his date themselves having grandchildren. Although it is certainly possible that such a contingency will occur, its probability is not usefully measurable.

Risk, on the other hand, in its archetypal form is the situation we face when gambling on the roll of a “fair” die or the toss of a “fair” coin. We do not know what the particular result of that toss will be; our money is at risk. But we do know, however, the long term odds

21. See infra Section II.D.

22. The authors on decision theory each tend to label these divisions slightly differently. Theorists like Raiffa tend to lump all lack of information into one category which he labels “uncertainty” in RAIFFA, supra note 8, at 104-27, and “risk” in R. DUNCAN LUCE & HOWARD RAIFFA, GAMES AND DECISIONS: INTRODUCTION AND CRITICAL SURVEY 13 (1957) [hereinafter GAMES]. To complicate matters, in the earlier work he calls by the name “uncertainty” what I refer to here as ignorance. Resnik’s very helpful philosophical work, MICHAEL D. RESNIK, CHOICES: AN INTRODUCTION TO DECISION THEORY 13-17 (1987), uses the terminology that I have adopted. See also SMITHSON, supra note 13, at 9 (offering the most extensive taxonomy, which names risk as a species of uncertainty).

23. RESNIK, supra note 22, at 13.

24. A “fair” coin, by definition, has an equal likelihood of falling on either side; the same holds to a “fair” die.

25. Actually there is no way of knowing, in the sense of a deductively certain rule, that the probability of an apparently “fair” coin is congruent with the observed long
of such an undertaking (1/6 for any particular side of the die, 1/2 for each side of the coin). Making probability judgments in these circumstances is not difficult. These judgments based upon known data are called frequency judgments. If, for example, the lawyer knew that in cases such as the plaintiffs', juries in her county had awarded $1,000,000 in 65% of all cases tried within the last two years, she would be justified on a frequency basis in concluding that she should assign a probability of .65 to the $1,000,000 verdict option instead of the .55 she assigned in the hypothetical. There are now services that collect and publish such data, but a purely frequency solution to problems of choice will never be totally available to lawyers. There will always be the problems of defining similar cases from similar jurisdictions with similar juries. In all but the very largest jurisdictions, sample sizes will almost always be too small for accuracy.

It is possible that people have not been watching long enough, although this seems unlikely considering the extent of human history with gambling, coins and dice. For discussion of the interrelationship of probability theory and gambling, with emphasis on the late development of the theory given the long prior history of gaming, see IAN HACKING, THE EMERGENCE OF PROBABILITY: A PHILOSOPHICAL STUDY OF EARLY IDEAS ABOUT PROBABILITY, INDUCTION AND STATISTICAL INFERENCE 1-11 (1975). It is also possible that the physical laws on which these observations depend will at some point in the future change, but if this were to occur, I would suggest that it would focus mankind's attention on issues other than the implications for decision theory. The only way to arrive at this result deductively is to a priori define a coin as "fair." That can happen, of course, only in theory. See discussion of induction and deduction infra Section III.A.

26. See JURY VERDICT RESEARCH, INC., PERSONAL INJURY VALUATION HANDBOOKS.

27. Sample size is an important determinant of accuracy. Many cases are necessary before samples reliably reflect the populations from which they are drawn. A one hundred fold increase in sample size will be necessary to obtain a ten fold increase in accuracy. See RUSSELL LANGLEY, PRACTICAL STATISTICS SIMPLY EXPLAINED 45-47 (Rev. ed. 1971) [hereinafter PRACTICAL STATISTICS]. People in general and even those with statistical training have a poor practical appreciation for the fact. RICHARD NISBETT & LEE ROSS, HUMAN INERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 77-82 (1980) [hereinafter HUMAN INERENCE]; Daniel Kahneman & Amos Tversky, Subjective Probability: A Judgment of Representativeness, 3 COGNITIVE PSYCHOLOGY 430-54 (1972), edited and reprinted in DANIEL KAHNEMAN ET AL., JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 38-46 (1982) [hereinafter HEURISTICS]. See also Daniel Kahneman & Amos Tversky, Belief in the Law of Small Numbers, 76 PSYCHOLOGICAL BULLETIN 105-10 (1971), reprinted in HEURISTICS, supra, at 23-31. Judgments based on small samples must be treated with extreme caution.

On the other hand, linear models (i.e., mechanical predictions based upon precise inputs) have outperformed ad hoc judgments in many prediction tasks; so some data, even if flawed, may be of value here. See Robyn M. Dawes & Bernard Corrigan, Linear Models in Decision Making, 81 PSYCHOLOGICAL BULLETIN 95-106 (1974); Robyn M. Dawes, The Robust Beauty of Improper Linear Models of Decision Making, in HEURISTICS, supra, at 391-407.
Until the 1930s, those advocating the use of decision theory only in instances in which frequency data were available were in the ascendancy, but at least since the publication of \textit{Truth and Probability} by Frank Ramsey, another school has emerged, and it has held sway in the field of decision theory since the 1950s. Subjectivists or Bayesians, as they are known, are disinclined to exclude problems from decision theory no matter how sparse the data upon which probability judgments are predicated. Their belief is that better decisions eventuate if, even in the face of what many would call ignorance, one makes a best guess at what the probabilities might be. The Bayesian position is that these \textit{statements} of probabilities are \textit{statements} about degrees of belief and therefore can legitimately be subjective in nature. Being taken with the rationality imposed by decision theory, they are loath to let its use gravel for want of matter. Further, through the use of Bayes's Theorem (discussed below), they maintain that original subjective judgments are subject to correction through the accumulation of subsequent data. If most lawyers are to use decision theory, their prior probabilities will almost always have to be of a Bayesian sort.

Since, for subjectivists, probabilities are expressions of degrees of belief, it will surely be possible for the lawyer to assign these numerical probabilities to the contingencies of the trial. When asked, as I suggested

Ross and Nisbett also indicate that sample size problems seldom have devastating impacts in real life modeling problems because even small samples tend to mirror the scatter characteristics of the population from which they are drawn. \textit{See} \textit{Human Inference}, \textit{supra}, at 256-60.


30. Named after the Reverend Thomas Bayes, an 18th century clergyman, whose treatise, \textit{An Essay Toward Solving a Problem in the Doctrine of Chances}, was sent posthumously to the Royal Society by his executor in 1763. This essay sets forth Bayes's theorem, which provides a method to adjust the prior probabilities in light of new data and then render in mathematical form one of the processes of induction. \textit{See infra} section II.E. For a historical sketch of Bayes by G.A. Barnard and a reprint of the essay, \textit{see} \textit{Press, supra} note 9, at 181-217.

31. \textit{See} \textit{Raiffa, supra} note 8, at 104-28, 154-56; \textit{see also} \textit{Games, supra} note 22, at 275-306.
“Where do you get these numbers?” a good Bayesian feels no compunction in saying, “Out of my head.” But what if the lawyer is less convinced and is reluctant to assign precise values? The Bayesians would assert that she must do so if she seeks to be rational.33 Bayesians would tolerate sensitivity testing of the result to changes in the assignments. Sensitivity testing involves making small changes in assigned values and noting the effect of those changes on the final result. In the example, if the lawyer did this sensitivity testing, she would find that if she altered all the assignments by decreasing the value of the most likely outcome by .05, the result would drop the expected overall outcome to $707,040. Since that value is lower than the settlement offer of $800,000, she might well want to rethink her assignments.35 But as the next section reveals, she must use these tests only to refine her assignments, not to avoid the assignment requirement.

D. Cover All Options

Is it possible to foresee all the contingencies? The second technical requirement of decision theory is that all assigned probabilities must be mutually exclusive and must sum to one.36 That is,

$$P(X_1) + P(X_2) + P(X_3) + \ldots P(X_n) = 1.$$ 

These requirements make sense when you realize that “one” is the equivalent of certainty in this system. If “one” represents certainty, then any combination of uncertain outcomes which together describe all possibilities inherent in the situation must by definition add to this certainty. It is a necessary artifact of the definitions of the system that this be so as long as the probabilities do not overlap.

There are three necessary, but not self-evident, corollaries which result from this definition. The first of these is that the probability of an

32. See supra text following note 14.

33. Raiffa discusses an example in which the evaluator has only a brief glimpse of physical objects and is, on the basis of this impressionistic information, asked to assign them by percentage to categories. He classifies this condition of very limited information as objective vagueness and asserts that one should assign probabilities as if one were dealing with event uncertainty. RAIFFA, supra note 8, at 104-08. Note also his further discussion of methods for dealing with this “vagueness.” Id. at 157-79. See YATES, supra note 8, at 31-32 for a pithy statement of the problem. See GAMES, supra note 22, at ch. 13.

34. See SMITH, supra note 9, at 52, 63; RAIFFA, supra note 8, at 140-41.

35. Compare with result supra text accompanying note 19.

36. See RAIFFA, supra note 8, at 110-16.
event and the probability that that event will not occur must also sum to one. Let me illustrate with an example in which the only issue is whether some event—e.g., granting of a directed verdict—will or will not occur. This is written:

\[ P(X) + P(\bar{X}) = 1. \]

Again this result follows from the definitions above; if an event and its complement are the only two possible results then one or the other of them must surely happen, and the sum of their likelihoods must be certainty. Few of us would be disturbed at all by this result in situations such as directed verdict or no directed verdict, because we are quite sure that \( P(X) \) and \( P(\bar{X}) \) really do exhaust the options. Either there will, at some point in the trial, be a directed verdict or there won’t. But what about situations in which we are not so sure what the options are? I think it helps one accept the Bayesian construct here to think of \( P(\bar{X}) \) not as “no directed verdict,” but rather as “none of the above.” This definition leaves open the option that something which you had not contemplated at all will occur. An example of such an unforeseen contingency might be loss of sufficient jurors from the venire during trial that a mistrial would take the case from the jury.

The utility of this definition appears more intuitively in the multiple option situation. For this purpose look at the damages node of the hypothetical. In describing the situation above, for the sake of simplicity, I had the lawyer assign probabilities to fixed dollar amounts which summed to one.\(^7\) This approach makes the math straightforward, as each probability can be multiplied by a fixed sum. But in the real world most lawyers would be disinclined to be certain that they could isolate all of the likely jury verdicts. The lawyer would almost certainly prefer to set aside some percentage of the available probability against the contingency that the jury might come out with some unexpected result; in effect to assign some probability to “none of the above.” The problem, of course, is how to determine the value of this probability. Since the value is by definition unknown, it can only be approximated. Various averaging techniques suggest themselves, such as taking the mean or median of the range within which the lawyer feels the unknown verdict

\(^7\) It should be evident that by having all probabilities sum to one, the system preserves the rule of complements adding to one which we identified above. \( P(X_1) + (P(X_2) + P(X_3) + \ldots + P(X_n)) = 1 \) is the same as \( P(X) + P(\bar{X}) = 1 \) because by definition since each probability is mutually exclusive \( (P(X_2) + P(X_3) + \ldots + P(X_n)) = P(\bar{X}) \).
Lawyer Decision Making

may fall. However, aside from the valuation problem, there is no major intuitive hurdle to cross in setting aside some fixed portion of the available probability to cover the contingency that some unexpected event will occur.

The third and thorniest corollary of the additivity requirement is that all kinds of unsureness about the likely results can properly be treated as if one were assigning probability to an unknown event, contingency, or risk. In the section above, the theory required the lawyer to assign precise probabilities. Here she is required to assign full weight to those probabilities. She is not permitted to say, "I think the liability likelihood is .85, but since it's only a hunch, I would like to assign .80 and leave .05 as an unassigned fudge factor." Or to take another kind of unsureness, if the case contained a prayer for injunctive relief the lawyer might say, "I am 95% sure that we will get the injunction, but the wording thereof will be critical to enforcement, and therefore, I won't be able to say whether we have real relief until I actually see it."

These various kinds of unsureness have been defined in the literature as "ambiguity," "vagueness," "fuzziness," etc. I will use the term "ambiguity" generally to define both these and other situations in which something other than the occurrence of the event is in question. Bayesians are quite insistent that cases of ambiguity should all be treated as if they were forms of risk which should be factored into the basic probability calculations. In the view of the Bayesians, the only issue

38. For example, if the lawyer were to conclude that there was a .05 chance of a verdict less than $500,000 but greater than $100,000, she might choose to assign this probability an overall value of $300,000, which is the arithmetic mean or average of the range $500,000 > Range > $100,000. The value of this probability then would be .05 * $300,000, or $15,000.

39. For a mathematical decision theorist who does permit the reservation of unassigned probability, see GLENN SHAFER, A MATHEMATICAL THEORY OF EVIDENCE 5-6 (1976) [hereinafter EVIDENCE].

40. Smithson gives the example of an evaluator who is quite sure of what he calls a "fuzzy" event. "Right now we have light to moderate drizzle." SMITHSON, supra note 13, at 94-99. He points out that advanced mathematical models are being developed in an attempt to deal with this "fuzziness." Id.

41. See SMITHSON, supra note 13, at 9, for detailed taxonomy of the general category that the author labels "uncertainty."

42. See RAIFFA, supra note 8, at 107:

In summary: The probability that the unknown urn is a $\theta_i$, denoted by $P(\theta_i)$, is clearly .8 in the original problem [when the observer knew the proportions], and it still ought to be .8 for Variation I [where only impressionistic information is available]; also, mathematically speaking, .8 is .8 is .8 is . . . , and it's of no avail to say that one .8 is fuzzier than another .8. See also D.V. Lindley, The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems, 2 STATISTICAL SCI. 17, 24 (1987):
is to get the decision maker to assign crisp probabilities in the face of the vagueness or ambiguity inherent in the situation. They offer a variety of methods, the roots of which involve comparison of the decision maker's vague probabilities to known lotteries.43

Two objections to this position are worth noting. First, where the effects of "ambiguity" on subjects have been tested, it is quite clear that people treat ambiguous problems differently from those involving the mere uncertainty or randomness.44 Secondly, new mathematical models which attempt to deal with these distinctions are under development. The one drawing the widest public attention at the moment is "fuzzy set" theory. That theory rejects the notion required by probability theory that options are mutually exclusive. The classic example is the class of "tall" people. A person 5'11" might have some membership in this class, but might also have membership as well in the class of "middle sized" people. Many qualities, including the effectiveness of an injunction, are hard to define in static categories. By permitting these flexible assignments and using a variety of mathematical techniques to manipulate the resulting groupings, "fuzzy set" theory attempts to model more closely the realities of natural language use and human thought processes.45 It is beyond my task here to value the claims of these

43. For a description, see RAIFFA, supra note 8, at 108-10, 161-65; see also SMITH, supra note 9, at 41-61.

44. The clearest example is the "Ellsberg Paradox." For Ellsberg's original discussion, see Ellsberg, supra note 28, at 89-112. Essentially, people asked to assign a probability to an ambiguous choice between two lotteries will assign a .5 probability to each in the absence of any information. This is consistent with the "principle of insufficient reason," which captures the intuitive sense that when people have no basis to discriminate between two selections, they will, and some say should, assign equal likelihoods to each. RESNIK, supra note 22, at 35-37. But when given a choice between these lotteries and one with a known probability of .5, they uniformly choose the known value. Ellsberg contends that people making these judgments are behaving rationally to protect themselves against unknown downside risks. Ellsberg, supra note 28, at 104-06. For examples of ambiguity effects in the selection of medical treatments, see Shawn P. Curley et al., An Investigation of Patients' Reactions to Therapeutic Uncertainty, 4 MED. DECISION MAKING 501 (1984); Shawn P. Curley et al., Psychological Sources of Ambiguity Avoidance, 38 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 230 (1986).

45. On fuzzy set theory, see BART KOSKO, NEURAL NETWORKS AND FUZZY SYSTEMS (1992); L.A. Zadeh, Fuzzy Sets, 8 INFO. CONTROL, 338 (1965) (by the father of fuzzy set theory). For a brief overview, see SMITHSON, supra note 13, 108-18. Attention also needs to be paid to EVIDENCE, supra note 39.
models to achieve the descriptive power that they assert, but rejection of these models in their early stages seems short-sighted.

All of these alternate models share the property of being sub-additive with respect to probability assignments. That is, the probabilities assigned to known contingencies do not sum to "one." Again, given their premises this makes sense. If not all unsureness is defined by event uncertainty or risk, then certainty or "one" is not to be found simply by adding together these event probabilities. Some measure of uncertainty must be left for the other sources of ambiguity. See Figure 6. (Black areas indicate unambiguous probability; grey areas indicate areas that sub-additive theorist would leave unassigned but Bayesians require be included in probability assignments.) Acceptance of the decision theory model, which normatively requires that all uncertainty be grouped together, explicitly rejects these alternate views.

Figure 6

46. However, the first generally accessible fuzzy set software shell has just been released, so study of those claims may not be that far away. See CubiCalc by HyperLogic.

47. Smithson suggests the possible emergence of a paradigm shift in decision theory. SMITHSON, supra note 13, at 145-51.

48. See, e.g., SMITHSON, supra note 13, at 120, 124, 128.
E. Reduction through Combination

Is there a satisfactory way to combine probabilities? In the example, the lawyer wanted to know the combined effect of getting to the jury, winning on the issue of liability, and achieving a particular verdict. The final assumption of the basic decision theory model is that the combination of probabilities is multiplicative. This assumption needs a little translation. As the example above showed, the lawyer predicted that there would be no directed verdict \( P(J_{nv}) = .95 \) and that then there would be a verdict in favor of her client on liability \( P(L_{liability}) = .85 \). If an event cannot occur unless another precedes or accompanies it, then that event is "dependent." For the case to reach the point at which damages are determined, both of these acts are required: the independent ruling of no directed verdict and the dependent finding of liability. The lawyer needs to be able to determine what the chances are that both will occur. There must be some way to combine them and figure out the joint probability.

The easiest way to see the appropriate answer is to use what is called a Venn diagram.\(^{49}\) See Figure 7. The lawyer is interested in the area in which the two events overlap, the area shaded in gray. It is evident from this diagram that the area of overlap is smaller than or equal to that occupied by the smaller of the two events.\(^{50}\) The chance of having no

\(^{49}\) Such diagrams are named for their developer, John Venn, an English logician who lived from 1834 to 1923.

\(^{50}\) People routinely behave, in some situations, as if joint probabilities are in fact larger than they are. See, for example, one of a series of studies on this issue, by Tversky and Kahneman in which subjects in 1980 were asked to rank order the likelihood of four events or combinations of events.

Suppose Bjorn Borg reaches the Wimbledon finals in 1981. Please rank order the following outcomes from most to least likely.  
1. Borg will win the match.  
2. Borg will lose the first set  
3. Borg will win the first set, but lose the match.  
4. Borg will lose the first set, but win the match.  
The answers they received rated option one (1) as the most likely followed by option four (4), option two (2) and then three (3). The rating of option four (4) as more likely than option two (2) violates the rule that compound probabilities must be smaller than or equal to the smallest of the probabilities being combined. Tversky and Kahneman ascribe this effect to the construction of imaginative scenarios which are then used to predict outcomes. Example subjects found it easier to envision a scenario which involved Borg's ultimate triumph rather than one that involved a simple loss. Amos Tversky & Daniel Kahneman, Judgments of and by Representativeness, in HEURISTICS, supra note 27, at 84, 96. The authors of HUMAN INFERENCE, supra note 27, at 115-122, link this scenario building with causal attribution, which raises issues that I will discuss below. See infra Part IV.
directed verdict and a finding of liability cannot be larger than the chance of having a finding of liability alone. Therefore, what is needed is a system that will account for the uncertainty of the two figures and will result in a smaller product than would be produced by the jury verdict area alone. It is a property of the multiplication of numbers less than one that the product is a number smaller than those multiplied. Here the joint probability of both getting to the jury and having the jury find liability is $P(J \cap L) = P(J) \times P(L | J)$. In the example, $.95 \times .85 = .8075$.

This property works very nicely in terms of describing joint probabilities, but it has an unfortunate side effect for the legal use of probability theory. As pointed out by L. Jonathan Cohen in *The Probable and the Provable*, application of this principle renders it all but impossible to prove a complex case. How can a mere arithmetic manipulation have such untoward effect?

51. L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 58-67 (1977). As with many of Cohen's points, he states his legal premises a little more dogmatically than is justified, but his fundamental complaint here is sound. He seems to maintain that the number of issues in a case should not have any effect on the difficulty of proof, which, as a practical matter, is to claim too much. But he is surely right that the number of issues should not make proof impossible.
To reach a conclusion of liability in a civil case or for guilt in a criminal matter, the fact finder must conclude that each element of the various causes of action is proved to the requisite degree of legal certainty (preponderance, beyond a reasonable doubt, etc.). Further, Cohen argues persuasively that the entire case must be in some way convincing. But let me assume that the plaintiff proves each element in even a simple negligence case by a probability of .52. Thus, the plaintiff convinces the jury that the defendant owed a duty of care ($P(D_{uty}) = .52$), that the defendant breached that duty ($P(B_{reach}) = .52$), that the breach caused injury to the plaintiff ($P(C_{injury}) = .52$) and that the plaintiff was damaged thereby ($P($damage$) = .52$). The jury's job under decision theory is now to determine the joint probabilities to determine whether the plaintiff has proved his case on liability.

But we know from the rule requiring multiplication that the joint probability here would be $P(D_{uty}) * P(B_{reach}) * P(C_{injury}) * P($damage$) = P(Liability)$. And $.52 * .52 * .52 * .52 = .073$. Therefore, although the plaintiff has proved each element by more than a preponderance, he has, under the rules of conjunction, totally failed to state a case which should convince the jury overall. If one may freely assume higher quanta of persuasion on each element in the typical case so as to push the overall probability higher, one must also allow that many law suits involve far more elements than the simple negligence case which I have laid out above. The result is that if our lawyer is to use probability theory in assessing the likely case outcome, and if she assumes that the jury will use the same reasoning process which she employs herself, she will inform her client that his prospects are bleak. As I will discuss below in the sections on inductive probability

52. COHEN, supra note 51, at 66-67.
53. The combined probability of .073 is premised on the assumption that all of the element probabilities are independent. When this is the case, a special statement of the conjunction rule holds that $P(X_1) * P(X_2) * ... * P(X_n) = P(X_1 & X_2 & ... & X_n)$. This produces the extreme result here. But even if one assumes that duty and breach are dependent variables, we will have problems. Using the formula for dependent probabilities, leaving damages as independent, and finally assuming the most extreme case in which, if duty is proven, breach and damages are certain, the following results: $P(D_{uty} & B_{reach} & C_{injury} & $damage$) = P(D_{uty}) * P(B_{reach} | D_{uty}) * P(C_{injury} | B_{reach} & D_{uty}) * P($damage$) or $P(D_{uty} & B_{reach} & C_{injury} & $damage$) = .52 * 1 * 1 * .52 = .2704$.
55. Nor is this concern purely hypothetical. It appeared in reverse in the case of People v. Collins, 438 P. 2d 33 (Cal. 1968). In Collins, the prosecutor wished to bolster his eyewitness testimony as to the identities of two daylight robbers. An elderly woman had been pushed to the ground in an alley and her purse was stolen. She identified her attacker as a younger white woman with medium blonde hair. A bystander testified that, after he heard shouts from the alley, he saw a woman, white with a dark blonde ponytail,
run from the alley to a yellow car driven by a black man with a beard and mustache. The defendant and his wife who generally answered the description were arrested and brought to trial. The identification testimony was to some extent undermined on cross.

The prosecutor at that point put a professor of mathematics on the stand. The professor explained the rule of combination through multiplication. The prosecutor, one is tempted to say in the best Bayesian manner, provided numerical probabilities for the various pieces of the description provided:

- Yellow car: .1
- Man with mustache: .25
- Woman with ponytail: .1
- Woman with blonde hair: .33
- Negro man with beard: .1
- Inter-racial couple in car: .001

Multiplying these factors together, the prosecutor got an overall likelihood of a couple meeting these criteria as one in 12,000,000. He then argued that this probability was so low that the defendant, who met all of these requirements, must be guilty. Collins was convicted.

The California Supreme Court, in overturning the conviction, lodged four basic objections. First, they showed in an appendix to the opinion that, on the prosecutor's own figures, there were likely to be one or two other couples meeting the criteria in the metropolitan area in which the crime occurred. They held that it was not a sufficient ground to convict that the field of potential defendants had been narrowed to a small group of whom the defendant was one. Next, the court pointed out the lack of any factual basis for the prosecutor's probability assignments. They questioned the assumption that all of the probabilities were independent, as the prosecutor's calculations had assumed. Finally, the court worried that the jury would be dazzled by the formula and fail to realize that they still had the obligation to weigh the credibility of the eye witness testimony to establish whether the events had occurred as they asserted, i.e., to determine whether the various criteria used were in fact the correct ones.

No one would, I think, challenge that if the jury was prejudiced in the sense that they did not understand their continuing role as fact finder, then the evidence should have been excluded. So the fourth point in the opinion is not implicated. There are those who would contend that evidence of the size of the class of possible defendants should be admitted. But the points about probability assignments and independence are critical to us. On the first, all that the prosecutor was doing was assigning his degrees of belief to the various criteria. He invited the jury to substitute their own if they did not approve of his. He lacked frequency data, but as we have seen that would not stop a good Bayesian for long. Furthermore, it is not clear that his assignments were all that unreasonable. With the possible exceptions of that assigned to the occurrence of inter-racial couples, they do not seem unreasonable hunches to me. (Even if inter-racial couples were one in 100 (.01) the result of the multiplication would have been one couple in 1,000,000. If one in every ten couples were inter-racial the result would have been one in 100,000.) My guess is that the jury would have found either of these results significant.

On the issue of independence, the only two criteria which are certainly dependent to some degree are the man with the mustache and the negro male with a beard. Assuming that the "man with mustache" is totally dependent on the variable "negro with a beard", which is surely an unduly conservative assumption, the overall probabilities only change from one in 12,000,000 to three in 12,000,000 (or three in 1,000,000 or three in
and causal decision making, it is in part the rejection of this method of combining probabilities which underlies alternate methods of prediction.

There is a further byproduct of the rule of multiplicative conjunction. In addition to sanctioning the use of subjective probability judgments, the peculiarly Bayesian contribution to decision theory is the use of Bayes's Theorem itself.\(^5\) That theorem provides a means of adjusting a prior hypothesis or base rate information in light of new data. As such it is of high interest to legal decision makers because it permits the adjustment of likelihood judgments to accommodate new evidence. The theorem is a specialized application of the multiplication rules set forth above.\(^6\)

100,000, if the jury made the adjustments suggested in the paragraph above.) Again the resulting change is not likely to sway the weight given this evidence by the jury.

My point is not that the prosecutor was correct in his approach or that the court was wrong, but rather that cleaning up the math would not solve the root problem. It is inherent in the rule of multiplication that the number of the criteria unduly influence the result. Put together enough criteria—whether they be elements in a cause of action or features of an identification—and the probabilities shrink alarmingly. Had defense counsel been mathematically alert he would have countered by embracing the prosecutor's suggestion and proceeding as follows:

Members of the jury, the prosecutor has shown you that you should multiply your degrees of belief in the sub-elements of a proposition to determine the believability of the whole. Here the prosecution wants you to conclude that my client is guilty. To do that, he needs to convince you 1) that a crime occurred in this state, 2) that the indictment of the defendant occurred within the statute of limitations, 3) that that crime involved the taking of property, 4) from the person or control of another, 5) without the permission of the possessor, 6) that force or threat of force was used in the taking, 7) that the thing taken was of value, and 8) that the defendant did the taking (the classic elements of a robbery charge). If we assume that the prosecutor has convinced you of each of these points so that you would assign a .9 probability of each, all that remains is that we multiply them together. \(0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.43\). So even if you are convinced of each separate proposition, by the prosecutor's logic, you cannot be convinced of my client's guilt beyond a reasonable doubt.

56. See supra notes 29-31 and accompanying text.

57. Bayes's Theorem holds that the probability of a new hypothesis formed after the assimilation of new evidence should equal the ratio of the prior probability of the initial hypothesis multiplied by the probability of the new data given that prior hypothesis to the result of that calculation plus a complimentary calculation in which the probability of the contradictory initial hypothesis is multiplied by the probability of the new data given that contradictory hypothesis. This equation is written:

\[
P(H_{\text{new}} | \text{Data}) = \frac{P(\text{Data} | H_{\text{old}}) \times P(H_{\text{old}})}{(P(\text{Data} | H_{\text{old}}) \times P(H_{\text{old}})) + (P(\text{Data} | H_{\text{old}}) \times P(H_{\text{old}}))}
\]

The denominator can be understood as a statement of the impact of the initial hypothesis and its compliment on the total probability of the new data. To see this,
There is very substantial evidence that people do not use anything consult the Venn diagram below. For the most straightforward discussion of the theorem that I have encountered, including this Venn diagram, see YATES, supra note 8, at 134-36.

![Venn Diagram]

To illustrate its application, imagine a juror in a criminal case. Before the defendant takes the stand, this juror's prior probability judgment as to the defendant's guilt was .10. (See infra text accompanying note 62 as to why the prior probability must be greater than zero.) However, when the defendant takes the stand, the juror notes that he has "shifty eyes." Now this juror, as do many in our culture, believes he has good reason to conclude that the probability of "shifty eyes" given guilt is .60; that is 60% of guilty people have "shifty eyes." When pressed, our juror will concede that a few not guilty people also have "shifty eyes," but the juror believes that this fraction is small, perhaps .10. Given these data and these beliefs about the incidence of "shifty eyed" people, what the juror wants to know is how he should adjust his prior probability of guilt. Bayes's Theorem would produce the following conclusion:

\[
P(G_{ult}|SE_{yes}) = \frac{P(SE_{yes}|G_{ult}) \cdot P(G_{ult})}{(P(SE_{yes}|G_{ult}) \cdot P(G_{ult}) + P(SE_{yes}|G_{ult}) \cdot P(G_{ult}))}
\]

or

\[
P(G_{ult}|SE_{yes}) = \frac{.60 \cdot .10}{(.60 \cdot .10) + (.10 \cdot .90)}
\]

or

\[
P(G_{ult}|SE_{yes}) = \frac{.06}{.06 + .9}
\]

or

\[
P(G_{ult}|SE_{yes}) = .4 \text{ or } 40\%
\]
approximating Bayes's theorem in their every day calculations. And there is some evidence that if used on intricate problems its mathematics becomes unmanageable. But it does provide a consistent model for evaluating evidence and one which can be made accessible through the use of appropriate software.

However, L. Jonathan Cohen points out that it does have a significant flaw which normatively precludes its use by judges or juries in testing the hypothesis of guilt or liability. Because the theorem relies on multiplication, it requires that the initial hypothesis of guilt or liability be non-zero. The criminal law is the easiest case for the problem. There is a presumption in Anglo-American law that the defendant is innocent. This presumption continues throughout the trial and has its effect at the end when the prosecution must overcome this presumption to obtain a conviction. Therefore, we instruct the jury that they must begin the trial with the belief that the defendant is innocent. This would translate into an assignment of zero (0) probability to guilt. But if all other evidence in the trial is combined through repeated use of Bayes's theorem to update the hypothesis, the outcome

58. See, e.g., Daniel Kahneman & Amos Tversky, On the Psychology of Prediction, originally published in 80 PSYCHOL. REV. 237 (1973), and reprinted in HUERISTICS, supra note 27, at 48, 53-57; see also HUMAN INFERENCE, supra note 27, at 141-50, 156-60 (reporting on studies by the authors and by Kahneman & Tversky among others); YATES, supra note 8, at 208-10 (reporting other studies). Hogarth sets forth the classic blue cab problem in which subjects routinely neglect prior probabilities in the form of "base rate" data in favor of admittedly flawed eye witness testimony. HOGARTH, supra note 13, at 42-45. It is not at all clear to me that people are behaving contrary to rationality in making these choices, as the probability theorists contend. It seems much more likely to me that they simply don't know how to do the relevant calculations. But in any case, it is surely proven beyond question that for whatever reason, people don't use Bayes's Theorem or anything approaching it in everyday evaluations.

59. The assumption that all prior probabilities and conditional probabilities are or can be made reliably available is probably wrong. See YUN PEN & JAMES A. REGGIA, ABDUCTIVE INFERENCE MODELS FOR DIAGNOSTIC PROBLEM SOLVING 14-15 (1990) [hereinafter ABDUCTIVE]. See also SMITHSON, supra note 13, at 98-99, and Edward H. Shortliffe & Bruce G. Buchanan, A Model of Inexact Reasoning in Medicine, 23 MATHEMATICAL BIOSCIENCES 351 (1975).

60. See PRESS, supra note 9, at 85-100 for a partial list.


62. Let me hasten to say this is not a factual presumption, but rather an assignment of the burden of persuasion. See 2 JOHN WILLIAM STRONG ET AL., MCCORMICK ON EVIDENCE § 336 (4th ed. 1992). One intent on preserving Bayes's Theorem in this application might finesse the problem by saying one does not have to factor the presumption in with other proof, but to do so would clearly undermine what we think we are doing. If one does not use the presumption, from what basis should one begin the calculation?
will always be zero. Mathematically any number multiplied by zero is zero.

The fact that the theorem cannot be used by courts to determine guilt or liability does not, of course, preclude its use by courts or by lawyers for other purposes, but the counter-intuitive nature of some of these conclusions serve, at least, to give me pause.

F. Summary

Decision theory is designed to evaluate rival courses of action. It relies on the ability of its practitioners to assign discrete probabilities to events. It insists that those probabilities describe all possible contingencies. That is, it requires that the probabilities sum to "one" which connotes total certainty in the probability calculus. Finally, it decrees that the probability of multiple events will be smaller than or at most equal to the probability of the least probable event in the combination.

I will show below that both of the other models of decision making—inductive probability and causal explanation—reject these core hypotheses. Neither of them requires precise probability assignments, thus admitting a degree of uncertainty (ambiguity, ignorance, vagueness, etc.) into judgments which the decision theorists find intolerable. Neither draws the comfortable conclusion that its system necessarily covers all contingencies. Finally, neither inductive probability nor causal explanation adopts decision theory's handling of multiple events. As I explicate those theories below, the significance of these disagreements will emerge starkly.

III. Inductive Probability

A. Induction Defined

Most judgments under decision theory are inductive in nature.63 That is, the practitioner reasons from the characteristics of the available information to some conclusion about the underlying reality which the data reflect. The Concise Oxford Dictionary defines induction in logic as, "the inference of a general law from particular instances."64 Perhaps the

63. The exceptions come in games of chance with "fair" coins or dice. In those instances, it is possible to work out the odds a priori from the definitions of the game. See L. JONATHAN COHEN, AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY 15 (1989) [hereinafter PHILOSOPHY OF INDUCTION].

64. THE CONCISE OXFORD DICTIONARY OF CURRENT ENGLISH 603 (1990) [hereinafter OXFORD].
most recent memorable appeal to induction occurred during the 1989 confirmation hearings of Richard Darman as Office of Management and Budget Director. Mr. Darman advanced the test to be used to determine whether a proposed scheme was a tax which the administration would oppose. He said, "If it looks like a duck, walks like a duck and quacks like a duck, it is a duck."65

All induction proceeds on this sort of premise; evidence from the environment reflects more or less clearly the underlying laws which control that environment. One attempts to extrapolate from that evidence to discern the laws.66 Classic decision theory is inductive because the decision maker uses the information available (be that information frequency data or whatever contributes to her hunch)67 to predict likely states of the world at some future time.68

For much of the last twenty years,69 inductive thinking in decision making was synonymous with probability based decision theory: the method discussed above. Recently, competitive notions of choice and decision have begun to surface. I will discuss these notions in this section under the heading of inductive probability. It is particularly interesting to me that these notions are coming from two very different sources: philosophers of science and cognitive psychologists engaged in the study of artificial intelligence.70

First, briefly, I will make explicit the critique of decision theory contained in this model. Then I will explore at more length the

67. Under some decision approaches, the decision maker may not be conscious of what factors influences her hunch. Her thought process is a "black box." See Isaac Levi, Self Profile by Isaac Levi, in HENRY E. KYBERG, JR. & ISAAC LEVI 181, 202 (Radu J. Bogdan ed., 1982) [hereinafter Levi Portrait]. A decision maker may be able to identify some or all of the influences affecting her choice, but the failure to identify these influences does not automatically discredit her choice. Adopting this approach makes accepting a choice easier, but also robs it of its justification for others.
68. Unknown states of nature—i.e., how the relevant piece of the world is or will be at some relevant time be—are referred to in decision theory as θ, theta.
69. As Ian Hacking points out in his fascinating intellectual history of the early developments of probability theory, this dominance is a recurrent pattern in decision thinking. HACKING, supra note 25, at 14-16.
70. For example, see Levi Portrait, supra note 67; HOLLAND ET AL., supra note 66; PHILOSOPHY OF INDUCTION, supra note 63; ISAAC LEVI, THE ENTERPRISE OF KNOWLEDGE (1980) [hereinafter LEVI: KNOWLEDGE]; L. JONATHAN COHEN & MARY HESSE, APPLICATIONS OF INDUCTIVE LOGIC (1980) [hereinafter APPLICATIONS CONFERENCE]; and JOHN R. ANDERSON, COGNITIVE PSYCHOLOGY AND ITS IMPLICATIONS (3d ed. 1990) [hereinafter COGNITIVE PSYCHOLOGY].
contribution of inductive thinking to extrapolation from prior experience to solutions for new sorts of problems.

B. The Inductive or "Baconian" Critique of Decision Theory

L. Jonathan Cohen points out in The Probable and the Provable\(^71\) and Hacking established historically\(^72\) that it is not necessary to use the tools appropriate to the analysis of games of chance in the study of less structured, real world uncertainty.\(^73\) Cohen, who labels his inductive methods "Baconian" after Francis Bacon, is not the only theorist to raise these issues,\(^74\) but I will use his work to illustrate the point. It is true that use of these tools when faced with uncertainty provides a precise and rational predictive method, but it is Cohen's premise that this rationality comes at too high a price. It requires that the practitioner claim a degree of assurance which she may not possess,\(^75\) and it deflates the probability of the conjunction of independent events.\(^76\) Cohen is concerned with these problems, not in the abstract, but because he believes that they make decision theory a poor tool for either scientific development or legal analysis.\(^77\)

It helps to see the argument if I mix the critique with the alternative. Cohen's objection to the additivity requirement of decision theory is that

\(^71\) COHEN, supra note 51, at 2-3.
\(^72\) HACKING, supra note 25, at 11-17.
\(^73\) Hacking lays out the interconnections of people and ideas which initially gave rise to the theory. HACKING, supra note 25, at 57-72. Cohen calls these tools "Pascalian probability" after Blaise Pascal, the theorist generally credited with the original development of the probability calculus. Cohen refers to his own system as "Baconian" after Francis Bacon, the English lawyer and inductive philosopher. COHEN, supra note 51, at 43.
\(^74\) For example, see LEVI: KNOWLEDGE, supra note 70, at 89-98; Levi Portrait, supra note 67, at 202-05; and EVIDENCE, supra note 39.
\(^75\) COHEN, supra note 51, at 47, 81, 310-12; Isaac Levi captures the problem in his essay Potential Surprise: Its Role in Inference and Decision Making [hereinafter Potential Surprise], in APPLICATIONS CONFERENCE, supra note 70, at 3:

In presystematic discourse, to say that X believes that \(h\) to a positive degree is to assert that X believes that \(h\). To claim that X disbelieves that \(h\) to a positive degree implies that X disbelieves that \(h\) (i.e., believes \(\neg h\)). Finally, when X believes that \(h\) to a 0 degree and disbelieves that \(h\) to a 0 degree, he suspends judgement [sic] as to the truth of \(h\).

The problem with Bayesian probability is that there is no way adequately to express suspended judgment because of the requirement that probabilities be additive and complimentary. See discussion supra Sections II.C & D.
\(^76\) See discussion supra Section II.E.
\(^77\) COHEN, supra note 51, at 49-120; PHILOSOPHY OF INDUCTION, supra note 63, at 1-12.
it uses the wrong definition of uncertainty. In the closed universe of mathematical probability, if one does not totally believe in the truth of a proposition P, one perforce assigns that uncertainty to a belief in the converse proposition \( \neg P \). Thus if one were so unwise as to say, “I only believe in leprechauns .1,” one has effectively said, “I believe in no leprechauns .9.” Cohen points out another option. One could say, “I lack sufficient information to believe in leprechauns more than .1, but for the same reason I lack sufficient information to believe in no leprechauns .1 also. I just don’t know.” So while in mathematical probability zero (0) equals absolute disbelief in a proposition and conversely absolute belief in its negation, in Cohen’s system zero (0) means insufficient information to believe anything, either the proposition or its converse.

It is easy to see how this judgment fits much more easily into scientific investigation and legal analysis. To say that one does not have enough information to evaluate the theory of relativity is not to say that one disbelieves that theory entirely. To say that one begins a civil trial with an open mind, ascribing zero belief to either side, is possible under Cohen’s system but not in mathematical probability—a seemingly modest proposal. But once one gives up the closed universe of mathematical probability in which one believes either in proposition P or “none of the above,” the entire decision calculus flies apart. For it is no longer possible to predict outcomes when the total probability assigned at any level of the decision tree need not sum to one. There will always be

---

78. Cohen is not so frivolous as to use leprechauns in his example, but you get the point.

79. Among the non-Bayesians, there is debate about whether lack of acceptance should be described as zero (0) belief as Isaac Levi sets forth in Potential Surprise, supra note 75, at 3, or whether, as Glenn Shafer holds in EVIDENCE, supra note 39, at 84-85, a belief that is not accepted may nevertheless accrue positive support. Cohen straddles these two positions. He allows a proposition that subsequently has been falsified by a later test (and thus presumably not accepted, at least in the later instance) to nevertheless retain its support from previous tests that it has passed, but his negation principle decrees that, on the most recent test, if support for the hypothesis on the present evidence is positive (\( s[H \mid E] > 0/n \)), then support for the complement (\( \neg H \)) is driven to zero (0) and thus (\( s[\neg H \mid E] = 0/n \)). COHEN, supra note 51, at 133, 177. This result is not quite as perplexing as it seems. The distinction lies between managing notions of past and present support.

80. A state of equipoise in mathematical probability is expressed by assigning an hypothesis and its negation each a probability of .5, so it might be possible (though counter-intuitive) to finesse the situation of equal balance in this matter. See CMG Ockelton, The Use of Mathematical Probability in Assessing Corroborative or Convergent Testimonies, 24 RATIO, 61, 61-65 (1982). But that solution still leaves unaddressed the artificial requirement of mathematical probability that a decrease in belief in H is necessarily an increase in the belief that \( \neg H \) is true. One might easily doubt both propositions for lack of evidence and this state is inexpressible in mathematical probability.
some unknown level of unsureness built into any evaluation of the available options.

Cohen also insists that one give up the notion that the value of a conjunction is found by multiplying together the probability of all its elements. Rather he substitutes the rule that a conjunction is equal to the probability of the member with the smallest likelihood. He defends this decision by pointing out that if one grants belief because of the probability of an hypothesis, then one ought to grant to a more elaborate (conjoined) hypothesis at least the degree of belief one is prepared to give its weakest link.

Cohen needs this proposition because it obviates the problems which he identifies, and I discussed above, in which the combination, for example, of elements in a legal case artificially depresses the belief which the trier of fact is able to assign to the case as a whole.

So at Cohen's insistence, inductive reasoning is to be stripped of two of the important dictates of mathematical probability. One can no longer sum all probabilities to the certainty represented by "one." And combinations of probabilities may no longer be smaller than either of the two probabilities combined. The choice of theories is stark and

---

81. See discussion supra Section II.E.
82. COHEN, supra note 51, at 169, 221.
83. See supra note 55.
84. In the face of this evisceration of classical decision theory, is it possible to use these concepts in anything resembling the evaluation that was the topic of the section above? Well, yes and no.

One way to look at this problem is to say that our lawyer still knows how decision theory would evaluate her clients' options, she is just no longer sure of her evaluations of the odds. She could, nevertheless, chart what the outcomes would be under various probability assumptions and then use some strategy to pick her best option given her uncertainty. For a technical description of how this might be accomplished, see, Potential Surprise, supra note 75, at 17-24. Should she elect to do this, non-Bayesian decision theorists offer her a variety of tools. See RESNIK, supra note 22, at 21-45. Recall that decision theorists recognize a state, which I referred to as ignorance, in which probabilities cannot be effectively assigned. The lawyer is now in a similar state because although she can assign probabilities, her assurance that she is including all the risks of the situation has been removed. Therefore, she is not sure the probabilities she assigns are correct. Theorists disagree about whether she should choose the option that will give her a maximum value if the worst case happens (known as the maximin rule): here, certainly taking the settlement. (See Potential Surprise, supra note 75, at 23. Ellsberg indicates that decision makers in fact adopt a like strategy. ELLSBERG, supra note 28, at 110-11. Patrick Maher argues, however, that subjects who make the Allais and Ellsberg choices make other choices that are inconsistent with Levi's theory. Patrick Maher, Levi on the Allais and Ellsberg Paradoxes, 5 Econ. & Phil., 69, 77-78 (1989). In an accompanying response, Levi denies that Maher's findings undermine the usefulness of his approach in the circumstances of the Ellsberg paradox.) Or she could seek to minimize the regret her clients would feel if they gave up a good chance at trial (known
fundamental. Do you accept Bayesian probability with its insistence on the assurance of a closed universe in which you either believe in leprechauns or disbelieve? Or do you accept the notion that there is some threshold below which evidence is so scanty that neither belief nor unbelief are appropriate? If you fall into the Bayesian camp you are rewarded with a workable methodology for evaluating problems under all kinds of uncertainty, which has only one really anomalous feature in that joint probabilities seem artificially deflated. If you prefer to retain your agnosticism and favor a calculus like that of Cohen, you must reconcile yourself to the fact that your judgments will be far less precise and in some cases inconsistent.

However, the measure of the success of this alternate inductive theory is not chiefly whether it performs the evaluation function of classical decision theory well.\superscript{85} It is in solving another sort of decision problem that inductive probability comes into its own.

C. Extrapolation

Many problems with which a lawyer deals lack the structure of relatively clear options, which is the forte of classical decision theory. The lawyer in grappling with these problems must extrapolate some coherent vision of the future and make predictions based on this extrapolation. To extrapolate in this sense is to “infer more widely from a limited range of known facts.”\superscript{86} It is the activity by which the lawyer takes her base of experience and learning and uses it to structure and draw conclusions about a previously unencountered problem. Inductive philosophers, such as Cohen and Paul Thagard, and cognitive

\begin{itemize}
\item as the minimax rule: here, perhaps arguing for taking a chance on trial. (I say perhaps because the trial option is so close to the settlement here, and you will recall the lawyer is now so unsure of her judgments, that it is entirely possible that the settlement might in many instances minimize regret). Or she could adopt some variant of these. For a good clear, non-mathematical discussion of these options, see Resnik, supra note 22. There are defects with all these rules in that they lead to inconsistent decisions in some circumstances, which is why the Bayesians reject them. Other methods are under development which may provide a way out of the morass. See supra text accompanying notes 44-46.
\item Cohen nowhere discusses in his books the classic decision problem of the kind that I posited above. He even concedes in Philosophy of Induction that “Pascalian” probability may have its uses. Philosophy of Induction, supra note 63, at 211.
\item Oxford, supra note 64, at 415.
\item Cognitive psychology is the term self-consciously adopted to distinguish what was, in the late 1950s, a renegade school at odds with the overwhelmingly behaviorist emphasis of the day. See Jerome Bruner, Acts of Meaning 2-11 (1990); Bernard J Bears, A Cognitive Theory of Consciousness 7-13 (1988); Cognitive Psychology, supra note 70, at 7-10. Behaviorists rejected the notion that one could
\end{itemize}
psychologists, such as John Holland, Keith Holyoak and Richard Nisbett, offer strategies for problem solution which can readily be turned to assist the lawyer in this work.

First, let me hypothecate the kind of problem involved here. The lawyer whom I introduced earlier confronts another decision problem. This time she is interviewing a prospective client and is trying to decide whether to take the case. I will suppose that the prospective client was injured on the job and is already receiving worker’s compensation. I will further suppose that the state in which the lawyer practices sets a very high burden of going forward in order to reach the jury in a case where damages over and above worker’s compensation are being sought.

study consciousness, preferring to treat the organism under review—either rat or man—as a “black box” whose adaptation to the environment in terms of stimulus and response was the proper sphere of psychological study. Behaviorists adopted this position out of despair that an organism could provide any reliable data about the functioning of its consciousness. Cognitivists want to study what is going on inside the mind.

88. See Holland et al., supra note 66. This book actually contains the perspectives of both the cognitivists and the philosophers, as one of the contributors is Paul Thagard, a philosopher with a primary interest in computational modeling of the process of scientific discovery. The book has enjoyed general critical approval although some reviewers have objected to a gap that occurs in spots between theory and empirical evidence. See, Jeff Shragar, Book Review, 41 Artificial Intelligence 249 (1989); Andrew D. Cling, Rules, Models, and Inference: A Review, 16 Behaviorism 181 (1988); Diane F. Halpern, Book Review, 33 Contemp. Psychol. 437 (1988); K. J. Gilhooly, Book Review, 78 British J. Psychol. 563 (1987).

89. I am relying on authors who do not directly address the problem I wish to discuss. Cohen is concerned with describing scientific experimentation. Holland et al. advance a method by which people adjust generally to the world. I am applying their ideas in a context that they do not address. To the extent that that application does damage to those ideas, the fault is mine, not theirs. They also do not cite each other though both cite other inductivist thinkers, primarily those from the philosophical camp. There is some evidence that computer modeling of induction process has come to the attention of the philosophers. See Margaret A. Boden, Real World Reasoning, in Applications Conference, supra note 70, at 359-75.

90. Unlike the previous example, this is clearly the lawyer’s decision so ethical obligations to involve the prospective client do not arise.

91. In West Virginia, a provision like the one I have in mind here is found at W. Va. Code § 23-4-2 (c)(2) (1985):

The immunity from suit provided under this section and under section six-a [§23-2-6a], article two of this chapter, may be lost only if the employer or person against whom liability is asserted acted with “deliberate intention.”

This requirement may be satisfied only if:

(i) It is proved that such employer or person against whom liability is asserted acted with a consciously, subjectively and deliberately formed intention to produce the specific result of injury or death to an employee. This standard requires a showing of an actual, specific intent and may not be satisfied by allegation
this case, I will assume that the law requires the plaintiff to demonstrate that the employer knowingly and intentionally put the worker in danger, before his case will be allowed to go to the jury. The lawyer in our hypothetical must decide whether she should attempt to make this case (on a contingent fee) or whether she should decline representation. A Bayesian would tell her to use her best judgment to assign probabilities to the likely outcomes in the case—summary judgment, motion to dismiss, jury verdict, etc.—do her calculations and thus control her judgment. I think our lawyer might well be pardoned for saying that this advice is not helpful. She would complain that even if ultimately she might be in a position to make such a judgment, what she needs now is a methodology for determining what factors should be included in making her “best judgment.” How should she organize her understanding of the world so as to effectively address this problem?

The inductivists, whose views I will examine below, would suggest that she proceed roughly as follows. First she should consider what she knows about the context in which she is operating. She should consider the relevant law, the attitude of her likely judge, the character of her prospective client and opponent, her degree of skill and experience, and any other item in the environment which she believes will shape the decision. It is not possible to lay down non-context-sensitive rules for what these influences might be because each context, according to these theorists, involves different variables. Next she should attempt to isolate from these environmental characteristics as broad a range of variables as possible which she believes are relevant to her decision. The emphasis here is on the breadth or the spread of the list for reasons which will appear below. She may want to use various creativity enhancing tools at this point to try to broaden her list.

Now she has her list of variables, for example: recent state supreme court cases in the field, the political posture of her judge vis-a-vis employees and plaintiffs, the demeanor of the prospective client, the precise content of his testimony on this issue, the coherence of the story or proof of (A) Conduct which produces a result that was not specifically intended; (B) conduct which constitutes negligence, no matter how gross or aggravated; or (C) willful, wanton or reckless misconduct; . . .

92. In addition to the contingent fee issues, she must worry about Rule 11 of the Federal Rules of Civil Procedure, or its state rule equivalent, which bars the filing of frivolous claims.

93. See Potential Surprise, supra note 75.

Lawyer Decision Making

he tells, the time and money she can devote to discovery, etc. Next, she must test them mentally in progressively more complicated combinations with the intent of finding the circumstance or combination of circumstances that will prevent her from prosecuting this case successfully. She will make these judgments based upon what she has noticed about this case in comparison to other cases with which she has experience. She is looking for instances of co-variation between the variables in her case and those in other cases that were or were not successful. With respect to each variable, she will insist on forming a judgment that raises the likelihood of her conclusion above some threshold with which she feels comfortable. She may find a new kind of decision tree useful here. See Figure 8. It illustrates her thought process through the first two layers of analysis. At each layer there are three options: She can fail to test her belief that the case is tenable (represented by the long, spindly branches on the tree); she may test the proposition

![Figure 8](image)

95. At any point that the lawyer neglects to examine succeeding variables, she can conclude that her case is secure, but in doing so she is relying on a weak reed. She does not know with certainty whether the hypothesis would have survived the tests avoided. This state of affairs occurs most frequently, not when the lawyer deliberately declines to test the thesis, but rather when she does not realize that additional constraints ought to be examined.
and find that it passes the test (here illustrated by the successful right fork at test (t1)); or she may test the proposition and find it wanting (here shown by the left at (t2)).

If she hits a variable which she believes will prevent successful prosecution of the case, she should examine the combination of variables that she has constructed to see whether changing it in some believable way will let her achieve that goal. Look in more detail at the branches from (t2). See Figure 9. If the facts do not fit because the client discloses items that are at odds with the needed elements (for example, if his own negligence or misfeasance caused the injury), then the way is blocked. But, for example, if the prospective client simply does not know the state of mind of the employer, the variable tested can be reformulated to permit further exploration. See test T2', Figure 9.

Let me highlight the five critical pieces of this description: 1) context sensitivity, 2) generation of a broad spread of relevant variables, 3) exposure to tests with progressively greater falsifying power, 4) the use of co-variation data, and 5) the requirement that each point reaches an

---

96. See COHEN, supra note 51, at 205-07, for the decision tree schema, although Cohen does not put it to the purpose I suggest.
acceptance threshold before it is adopted. Each of these prescriptions finds its equivalent in Cohen and in Holland et al.

**D. Context Sensitivity**

Let me begin with context sensitivity. Philosophers of science going back at least as far as Leibniz97 have tried to devise methods for regularizing inductive reasoning with either the lofty goal of discovering underlying methods that will assure correct results in all fields98 or the more modest objective of minimizing inductive errors in the field under study. The popular current thinking is that the lofty goal is unattainable.99 This belief is premised on the notion that inductive judgments must be made in the context of the field under study. In each field there are different sorts of problems to which different kinds of evidence are germane. Isaac Levi in Gambling with Truth100 gives the example of a scientist noting apparent cloud formations on Mars. The scientist would ask the source of these apparent phenomena and would consider issues of whether they represented sand storms, some sort of vapor, or a malfunctioning telescope. The inquiry would not extend, however, to a re-exploration of the theory of gravity or the Copernican model of the universe. Although both of those theories are inductively based, they are also firmly entrenched in the field of astronomy. The local field would, at any point in time,101 treat the first set of questions as available for study and the second as settled. This division of inquiry into areas that are currently relevant and those that are not, is critical, for

---

97. See HACKING, supra note 25, at 134-42.
98. See, e.g., Richard C. Jeffrey, Carnap and Inductive Logic, in RUDOLF CARNAP, LOGICAL EMPIRICIST 327 (Jaakko Hintikke ed., 1975). Carnap tried to construct a universal function which would relate evidence to hypothesis and tell a decision maker what degree of belief or credence he ought to have in the hypothesis, given the evidence.
99. See Jeffrey, supra note 98, at 327; HACKING, supra note 25, at 134-42; Levi Portrait, supra note 67, at 188; Jonathan E. Adler, Criteria for a Good Inductive Logic, in APPLICATIONS CONFERENCE, supra note 70, at 390 n.43; COHEN, supra note 51, at 24-25; PHILOSOPHY OF INDUCTION, supra note 63, at 126-30; HOLLAND ET AL., supra note 66, at 6-7.
100. ISAAC LEVI, GAMBLING WITH TRUTH: AN ESSAY ON INDUCTION AND THE AIMS OF SCIENCE 3-6 (1967) [hereinafter GAMBLING].
101. Levi recognizes that what are accepted beliefs at one point may be rejected later if a “deeper curiosity” displaces old theories. GAMBLING, supra note 100, at 4-5. For the recognized discussion of these scientific revolutions, see THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS (2d ed. 1970).
as we shall see below, skepticism as to all issues probably makes any conclusion untenable.\textsuperscript{102}

Cohen and Holland et al., vigorously assert the context dependence of inductive logic. Cohen points to the fact that claims of general theories are not empirically testable.\textsuperscript{103} Holland et al. criticize the generalists' approach for ignoring "environmental and cognitive realities."\textsuperscript{104} Both assert the role of the problem context as dispositive of the questions to be asked to test an inductive theory.

For the purposes of our lawyer's problem, context sensitivity means that she should ask questions that her training and experience have taught her are relevant to the solution of her problem. She should ask about the state of the case law, the nature of the facts, the disposition of the forum, the availability of needed resources, etc. In terms of her problem solution, it will generally not be helpful for her to speculate on the wisdom of the advocacy model of fact finding or cosmic meaning of truth in relation to her prospective client's perceptions. This precept of context sensitivity is in tension with the next portion of the model, posing a broad range of relevant question.

\textit{E. Relevant Variables}

Cohen's "Theory of Relevant Variables," as he inevitably denominates it, is best explained in his most recent book \textit{The Philosophy of Induction and Probability}.\textsuperscript{105} The critical point which he makes is that the appropriate degree of belief to accord a proposition rests not only with the frequency with which it seems to be confirmed, but also with the variety of circumstances over which it applies. This concept of spread he attributes in inductive thought originally to John Maynard Keynes who postulated it as the weight of a proposition but did not develop the implications.\textsuperscript{106} As a statistical notion, of course, the idea that the validity of a prediction rests not only on the position of the mean, but also on the breadth of the standard deviation,\textsuperscript{107} is not new. Holland emphasizes it as well and further demonstrates that it is likely that people use expectations about the spread of variables in making inductive judgments.\textsuperscript{108}

\begin{itemize}
  \item \textsuperscript{102} \textit{See infra} Section III.I.
  \item \textsuperscript{103} \textit{Cohen, supra} note 51, at 25.
  \item \textsuperscript{104} \textit{Holland et al., supra} note 66, at 5-9.
  \item \textsuperscript{105} \textit{See Philosophy of Induction, supra} note 63, at 96-99.
  \item \textsuperscript{106} \textit{Id.} at 102-03.
  \item \textsuperscript{107} \textit{See Practical Statistics, supra} note 27, at 51-91.
  \item \textsuperscript{108} \textit{Holland et al., supra} note 66, at 185.
\end{itemize}
Let me illustrate an aspect of the problem first with an example drawn from Cohen.\textsuperscript{109} Suppose that I know that on average nineteen out of twenty sailing ships (in the days of the East India Clippers) returned safely from their voyages. If I represent Lloyd's of London, I might feel confident in charging a relatively small premium for an insurance policy against total loss. But suppose not all kinds of ships have the same success rate. If ships with rotten hulls average only fifty percent returns, and if the owners of those ships take disproportionate advantage of my low rates, I will go bankrupt fast. See Figure 10 for an illustration of this bi-nodal problem. In legal terms, if the average verdict in all auto injury cases is $500,000,\textsuperscript{110} as a plaintiff's attorney I could be seriously misled unless I know that the average verdict where the victim did not wear a seat belt was $50,000.\textsuperscript{111}

\textbf{Figure 10}

There is an additional problem pointed out by Holland et al.\textsuperscript{112} There may be considerable overlap between categories. In Figure 11,
assume curve 1 represents personal injury plaintiffs with no residual impairment at the time of trial while curve 2 represents those with such an impairment. If the curves are like those shown in Figure 11, this difference is a very good predictor of verdicts because the average verdict is much higher and there is very little overlap between the curves. But if instead the proper curves should look like Figure 12, the value of this distinction as a predictor all but disappears even though the averages remain the same. This is because large numbers of both kinds of cases fall in both the high and low recovery areas.

Cohen's real contribution on this point is in his insistence that the reliability of an inductive judgment will rest fundamentally on whether the decision maker has identified a sufficient breadth of variables. Since by definition we cannot know statistically the true spread of the variables involved, we need to protect against surprise by insuring to the extent that we can that no specialized pocket of data is lying undiscovered and that the apparent distinctions are really predictive.

The objective here is to prevent being blind-sided. If the lawyer in the hypothetical neglected to consider the attitude of her judge toward the granting of motions to dismiss, it would be clear to experienced lawyers that she had made a fundamental mistake in assessing the variables relevant to her case, but it is precisely this type of mistake that young lawyers and non-lawyers make all the time. They assume that there is no variable there; that the only relevant issue is what the law says the standard should be.

But, Holland et al. add an additional empirical finding to the mix. They establish that people in general carry about with them a model of the likely spread of variables within a particular sphere. In experiments that they report, subjects were asked what conclusions about the reliability of a small sample they would draw in different fields: chemistry, ornithology, and human physiology. The judgments about reliability

113. Again the data are mythical.
114. COHEN, supra note 51, at 135-44.
115. Unlike a "fair" game of chance, the true proportions of a population spread on some variable cannot be known a priori or with true certainty unless the entire population is polled. One cannot know the true proportion of registered Democrats without consulting the voter rolls, and even then we know that that figure is only weak inductive evidence for the likely results of the next election. Statistical techniques, including the computations of standard deviations within a sample, are aimed at making these inductive guesses better, and they work fairly well if the researcher has a true sample of the "relevant" population and if there are no multi-nodal anomalies in the population spread. But the kind of data on which they are based are not normally available to the practical decision maker. PRACTICAL STATISTICS, supra note 27.
116. HOLLAND ET AL., supra note 66, at 250-52, emphasizes the role of expertise in bringing models closer to statistically normative predictors.
Figure 11

Figure 12
differed sharply and consistently, depending on the field. Subjects believed that a small sample of a chemical element could be safely assumed to be very representative, while even a fairly large sample of persons of a certain body type from an ethnic group was unlikely to be representative of that group as a whole.\(^{117}\) If this sort of conclusion is as generalizable as the theorists believe, then we carry with us some in-built notions\(^{118}\) of how wide the spread of relevant variables in a field is likely to be.

If our judgments about the likely number of relevant issues is not only consistent but also accurate, there is no problem. In that

---

117. Id. at 241-43. The experiment they report was conducted by Nisbett, Krantz, Jepson, and Kunda and published in 1983:

- Subjects were first given information about several novel kinds of objects and then asked for their generalizations about the objects:
  - Imagine that you are an explorer who has landed on a little-known island in the Southeastern Pacific. You encounter several new animals, people, and objects. You observe the properties of your “samples” and you need to make guesses about how common these properties would be in other animals, people, or objects of the same type.
  - Suppose you encounter a new bird, the shreeble. It is blue in color. What percent of all shreebles on the island do you expect to be blue?
  - Why did you guess this percent?

- Subjects were also told that the shreeble was found to nest in a eucalyptus tree and were asked what percent of all shreebles they expected to nest in eucalyptus tree. The subjects then were told to imagine that they had encountered a member of the “Barratos” tribe. He was brown in color, and obese. Finally, they were told that they had encountered a sample of a rare element called “floridium.” It was stretched out to a filament and heated to a very high temperature. The subjects were asked what percent of all floridium they expected to conduct electricity and to burn with a green flame.

Id. On one sample above, subjects concluded that between 90 and 100% of all “floridium” would be conductive and burn with a green flame. They rated “Barratos” skin color as having the same degree of certainty. But they were willing to allocate only a 70 to 80% certainty to shreeble color and nesting habits. Finally, they believe that there was only a 30 to 40% chance that male “Barratos” would be obese. Thus it would appear that people have clear and distinct models for variability depending on the kind of thing being measured.

118. The problem, of course, is that these in-built notions, which themselves must come from inductive experience, may be wrong. The lawyer who is convinced that her judge will not give more than a fixed proportion of marital goods to a divorcing wife may be right in predicting the variability of the judge’s behavior. But if, for example, she does not realize that the presence of domestic violence would markedly increase the judge’s generosity, then her predictions about amounts formed in the “normal” case may prevent her from seeking what is attainable for her client in the case where violence is present.
circumstance, not only will lawyers consistently look at the same number of issues, but it will be the right number as well. If we cannot safely conclude that this is the case, the tension between context sensitivity and variable spread becomes a serious source of concern. That is because context sensitivity tells the lawyer to concentrate on the issues that people in her field see as relevant and the findings of Holland et al. indicate that that list may be pretty well fixed; however, Cohen’s emphasis on variable breadth counsels against undue optimism that the right variables are in the consensus mix.

To revert to the main example, suppose that the lawyer in the job injury case did not see law reform as an option worth considering; it was not on her list of relevant variables to inquire into the changes in the attitude of her supreme court on issues of workers’ remedies. If we further assume that some law reform development was possible, then the lawyer is doing herself and her perspective client a potential injustice by failing to see the issue. In my view, truly successful decision making contains a heavy dose of creative problem finding.

The tension between maintaining purposeful inquiry and avoiding tunnel vision is endemic in the twin injunctions of context and breadth.

F. Hypothesis Testing

Once the lawyer has done the best she can at identifying the relevant variables, how do we evaluate the impact of those variables on the issue? Here Cohen and Holland et al. use very different language to describe remarkably similar ideas.

Cohen, remember, has his background in the philosophy of science. He sees the inductive process in the mode of the controlled experiment. Therefore, he describes the approach as follows: Conduct a series of progressively more complicated tests (these may be thought experiments) combining at each stage a new variable so that after each new experiment the decision maker will know whether that variable, in combination with the others already tested, defeats or aids verification of the hypothesis. Translated into the terms of the lawyer’s problem, she would first review the cases from her state court to determine the legal standard applicable. Assuming that some set of facts would meet that legal standard, she has passed the first variable. Next she should probably take her prospective client’s version of the facts in his case and

---


120. COHEN, supra note 51, at 131-33.
match them to the legal requisites. This is the second set of variables. If there is no match (i.e., if the facts as the prospective client describes them do not fit the available legal theory), the lawyer's hypothesis that this is a tenable case has suffered a set back. If she does not believe she can alter the law or discover additional facts that her prospective client may not possess (i.e., if she cannot in some way modify the variable), she must accept the falsification of her hypothesis and conclude that the case cannot be won. If, on the other hand, the facts do fit the legal standard, she has passed the second test and should proceed to the next, which might be one of her available resources.

Holland et al. are concerned with the reality of how people think and, to some extent, with replicating at least a facsimile of those thought processes on the computer. They do not see the decision maker's activity as a series of controlled experiments, but rather as the application of a mental model formed from a series of default rules that will be applied until they are found to be untenable. Thus, they would describe the lawyer's thought process as a set of the following condition/action rules:

Rule 1: If a person has been injured and if that person has a cause of action, then that person is entitled to damages.

[hunt for matching rule]

Rule 2: A person has a cause of action if he was injured through the knowing and intentional act of his employer.

[match found "cause of action"]

[hunt for new match]

......

Rule N: If my prospective client says and if my prospective client is credible in saying that he was injured through the knowing and intentional act of his employer, then, all other things being equal, he was injured through the knowing and intentional act of his employer.

[match found "knowing and intentional act"]

[reality check, client does not say that]

121. See Holland et al., supra note 66, at 29-101.
Rule N + 1: If my prospective client says and if my prospective client is credible in saying that he does not know whether he was injured through the knowing and intentional act of his employer, but if I can prove through discovery that he was injured through the knowing and intentional act of his employer, then, all other things being equal, he was injured through the knowing and intentional act of his employer. 122

Under both the Cohen model and that of Holland et al., the lawyer would submit her hypothesis that the case was tenable to a series of tests or condition/action rules. Normally, if a particular test or rule endangered her hypothesis, she would attempt to alter that test or rule to find a way to preserve the hypothesis. 123 So, for example, when the prospective client’s testimony did not provide all of what was needed to meet the legal standard, she altered the test or rule to accommodate additional sources of proof. This alteration of the specific test or generation of a new rule gives the decision maker a great deal of flexibility and resilience. But both Cohen and Holland et al. allow that this resilience comes at a price.

In Cohen’s system, the resilience is termed tolerance for anomalies, 124 and for Holland et al. it is called preservation of defaults. 125 Both point out correctly that a decision maker who had to jettison her basic hypothesis or her underlying rules because of some peripheral conflict, would have a very brittle inductive system. Cohen

122. There are obviously some major assumptions built into this truncated model, not the least of which is “all other things being equal.” In real life, decision makers can make some relatively safe background assumptions about normal conditions: gravity, oxygen, the anglo-american jury system, etc. One of the values of thinking rules out in this stilted sort of pseudo-code (computer talk for this kind of writing) is that it pushes one to spell out those assumptions to avoid the kind of sloppy thinking evidenced here. In this case, “all other things being equal” means the prospective client had the opportunity and capacity to observe and report accurately what he saw. This is a rather large mouthful to take for granted.

123. Both Cohen and Holland et al. agree that the results of some tests are so disconfirming that an entire reappraisal of the problem, not just tinkering at the edges, is required to create an accurate model. See COHEN, supra note 51, at 179-80; HOLLAND ET AL., supra note 66, at 330-33. Although in their view the old model is less likely to be replaced than displaced by a narrower, but still existent, sphere of influence.

124. COHEN, supra note 51, at 162-66.

125. HOLLAND ET AL., supra note 66, at 183-85.
uses the example of Newtonian mechanics and the motion of the Moon. For a long time Newton's laws were thought to produce results at odds with the observations of the movements of that body. Yet no one rejected Newton's system with its extraordinary explanatory power because of this anomaly nor would it have made sense to do so. Ultimately, according to Cohen, the calculations were revised so that the system and the observation were harmonized. Likewise, it would make no sense for the lawyer to discard her understanding of the general relevance of a client's evidence just because in this circumstance that evidence is not helpful.

However, there are circumstances in which the edifice of theories does need to be shaken. Again according to Cohen, there was another anomaly unexplained by Newton's laws: the movement of the perihelion of Mercury. That anomaly was also disregarded by physics for 250 years until it was explained by Einstein's general theory of relativity. Now, in the case of this paradigm shift, I am not greatly disturbed because seventeenth century physics did not have the equipment or other theories that would have been necessary to appreciate Einstein in any case. I don't feel a great sense of loss in the delay. But the tolerance for anomalies can translate, as Holland et al. show, into a protection of unjustified beliefs which are inimical to the decision maker and perhaps to others. If for example the lawyer's rule structure called for implicit belief in everything her prospective clients told her, she would have a great deal of trouble making case acceptance decisions. If that default rule remains entrenched in her structure, significant numbers of anomalous experiences may not help to uproot it because those counter examples will just generate the formation of lower level special case defaults (e.g., I will believe all clients except those charged with misdemeanors or I will believe all clients except those with red hair). Amendment at the edges leaves a potentially fundamental flaw unaddressed.

126. COHEN, supra note 51, at 163. "It was not till 1752 that Clairaut showed how the theory could be made to produce results that agreed with the observed movements." Id.
127. The point of the orbit nearest to the Sun's center. See OXFORD, supra note 64, at 885.
128. COHEN, supra note 51, at 163.
129. The protection of old rules which are left in place because special exceptions are created to explain counter instances may explain why people cling to ethnic and racial stereotypes, despite an apparent real friendship for a specific member of the stereotyped group. See HOLLAND ET AL., supra note 66, at 221-22, 245-49.
G. Co-Variation

I have stylized the model by postponing to this point discussion of the role of co-variation. By co-variation I mean the realization by the decision maker that phenomena coincide. Without the recognition of co-variation effects, it would be impossible to make any predictions. If clouds do not, at least sometimes, presage rain, then noting an overcast sky outside my window gives me no help in deciding what the weather may be like in an hour. Similarly, if the fact that a client remembered certain facts at an in-office interview did not provide at least some assurance that that client would remember the same facts later on, planning a trial strategy would be impossible.

Where the stylization comes in is in the fact that co-variation not only helps evaluate the likely effects of test or rule application, it is also critical in formulating tests and rules to begin with. The whole notion of "relevant" variables calls for the decision maker noting some correlation between the variable and the truth or falsity of the hypothesis. If there is no co-variation, there is no relevance. Holland et al. would insist that observation of co-variation precedes rule formation and that its role in the testing model laid out above is to assist the decision maker in forming new rules.

The problem with this part of the prescription lies not with the injunction, but with the execution. There is much research to indicate that human beings are fairly poor judges of co-variation; they tend to be misled by their expectations. In addition, in the terminology of information theory, there is too much noise in our environment.

130. See infra Part IV (discussing an important subspecies of co-variation: causation). But for the moment, reference is to the more general co-occurrence without any necessary causal link.

131. See HOLLAND ET AL., supra note 66, at 79-81.

132. HUMAN INFERENCE, supra note 27, at 10 ("There is mounting evidence that people are extremely poor at performing such co-variation assessment tasks."); see also id. at 93-97, 97-101; Dennis L. Jennings et al., Informal Covariation Assessment: Data-Based versus Theory-Based Judgments, in HEURISTICS, supra note 27, 211-30; HOGARTH, supra note 13, at 114-30; YATES, supra note 8, at 163-83. These commentators are most concerned that people pay more attention to what their theories lead them to think they see than to what actually occurs.

133. See JEREMY CAMPBELL, GRAMMATICAL MAN: INFORMATION, ENTROPY, LANGUAGE AND LIFE 18 (1982). Noise is the non-information imparting aspect of the environment. For example, static on a radio or a politician's deliberate attempt to obfuscate. In this case it refers to the fact that people seldom get information from the social environment that is not clouded by totally irrelevant data. One particular argument may have swayed the judge, but to insure against reversal he may list several other, to him unimportant, pieces of supporting evidence or logic.
the language of psychology, we lack reliable feedback.\textsuperscript{134} Judgments about human interactions, of the kind that lawyers are asked to make, typically involve long periods between the initiating action and the supposed corollary. The art of drafting a complaint will not be seriously tested\textsuperscript{135} in all likelihood until at least a year later when the case comes to trial and the judge finds it useful or useless in understanding the issues presented. In the meantime, many other things will happen in the framing of the issues so that it will seldom be possible to get a clear correlation between good complaints and well-focused cases. The lawyer’s rules about what constitutes a well-drafted document could seldom be justified in terms of any strong claim to correlation testing.\textsuperscript{136}

\textit{H. Acceptance Threshold}

Finally, I come to the issue of threshold, which is after all where I started. Cohen is unwilling to let the Bayesians assume that either one believes leprechauns or one believes no leprechauns. He wants to set a threshold of agnosticism below which one is free to believe neither or

\begin{itemize}
  \item \textsuperscript{134} See Janet L. Kolodner & Robert L. Simpson, Jr., \textit{Problem Solving and Dynamic Memory}, in \textit{EXPERIENCE, MEMORY, AND REASONING} 99, 101 (Janet L. Kolodner & Christopher K. Riesbeck eds., 1986); Hillel J. Einhorn, \textit{Learning from Experience and Suboptimal Rules in Decision Making}, in \textit{HEURISTICS}, supra note 27, at 268-83.
  \item \textsuperscript{135} This is assuming the complaint is sufficient to withstand the \textit{pro forma} motion to dismiss for failure to state a claim under Rule 12(b)(6) of the Federal Rules of Civil Procedure.
  \item \textsuperscript{136} For a good test of correlation, data are needed not only on the number of instances in which the tested variable did or did not predict the outcome, but also on the number of instances occurring without the presence of the tested variable. Thus on the good complaint/good trial test, data are needed on the number of good trials and bad trials given good complaints, but also on the number of good and bad trials given bad complaints. These data are typically displayed in a 2x2 matrix or four cell diagram.
\end{itemize}

\begin{center}
\begin{tabular}{|l|c|c|c|}
\hline
 & Good Complaint & Bad Complaint & Total \\
\hline
Good Trial & 10 & 8 & 18 \\
Bad Trial & 5 & 4 & 9 \\
Total & 15 & 12 & 27 \\
\hline
\end{tabular}
\end{center}

People also tend to concentrate entirely on the good complaint column of this matrix and conclude from data like that shown here (purely fictitious) that good complaints yield good trials. If the true data were like those shown, however, this conclusion would be untenable because exactly the same proportion of good trials obtained in the bad complaint column as well.
Lawyer Decision Making

credit each a little bit. Therefore, his system is non-additive and non-complimentary. On this he and Holland et al. agree. The lawyer should only attach belief in the winning nature of the prospective client’s case to the extent that her assessment, after all relevant tests have been performed, rises above some threshold. She is not required to believe in that proposition affirmatively just because she does not reject it out of hand. I have discussed the strengths and weaknesses of this position at length above. The arguments will not improve through restatement here.

I. Summary

I have developed a prescription for extrapolation which has the following characteristics. It requires context sensitivity to focus on the problem at hand but attempts to balance too comfortable acceptance of conventional wisdom by requiring that hypothesis be tested against all relevant variables. It requires progressive testing, with the benefit that it is resilient to unimportant anomalies but as a result may disregard important falsifications as well. It calls for observation of co-variation as its primary method for both rule creation and rule testing, thus being appropriately empiric and responsive to the environment, but there is evidence that faith in human perception of co-variation may be misplaced. Finally, it permits us to be agnostic about unproven propositions and their converses, but it does so at the price of opening the universe to unknown forces and destroying our assurance in a tidy description.

In the face of these inherent weaknesses, it is no wonder that David Hume, in 1739, enunciated the view that induction could never provide a satisfactory basis for belief because humans can never be sure that the world they see today will obtain tomorrow. Other philosophers would extend this skepticism to encompass the belief that people cannot even know the present with any accuracy.

You will recall Dick Darman and the duck. Unfortunately for the inductive mode of reasoning, and despite Mr. Darman’s assurances, it is possible to appear from a suitable distance to walk like a duck, look like a duck, quack like a duck and still be, if Hans Christian Anderson is to

137. See supra Section III.B.
138. See HOLLAND ET AL., supra note 66, at 61-64.
139. HACKING, supra note 25, at 176-85.
140. See supra Section III.G for a discussion of the limits of human perception. See also PHILOSOPHY OF INDUCTION, supra note 63, at 176-87. An example of this limit on human observation is to be found in the Heisenberg principle, named after German physicist Werner Heisenberg. That principle states that the momentum and the position of a sub-atomic particle cannot be determined at the same time because the measurement necessary to determine one will disturb the calculation of the other.
be believed, a swan. The problem is that you are never sure that you have identified all the right characteristics in such a way as to permit certainty that your theoretical definitions of birds (or taxes) in fact mirror reality. If you’ve got the right characteristics, you may nevertheless harbor wrong beliefs about the way nature really is, and your observations cannot be counted on to correct those beliefs because of the structure of the beliefs themselves and the weakness of human perception. Inductive reasoning must always be reasoning under some degree of uncertainty. You can suppress that uncertainty by treating it as if it is not there, but you cannot remove it from reality so easily.

Despite these drawbacks, people cannot give up inductive reasoning. If I could not conclude that because my alarm clock has always gone off in the past, it is likely to go off tomorrow morning, I would have to stay up all night to avoid oversleeping. Yet that conclusion depends on various inductive beliefs about the function of electricity, the physics of sound waves, the physiology of the human ear, etc. If the lawyer in the problem could not conclude that the court system would deal in a predictable way with this case, she could never project its potential for success.

So the lawyer is caught in the bind: She knows that inductive reasoning is inherently faulty, but she cannot get along without it.

IV. CAUSAL REASONING

A. Introduction

Both decision theory and inductive reasoning aid in determining what will happen next. In this part, I will examine causal reasoning, which seeks to determine why future events will occur. Why ask why? Three professions or disciplines use causal reasoning to meet three rather different objectives. Scientists seek to determine why events occur as they do so that science may deepen and simplify its understanding of the basic principles that govern our world. Their motives are both elegance and utility: elegance of a concept pleasing to the intellect, and utility of understanding the rules by which the world functions so we can bend those rules to our own purposes.

From a slightly different perspective, the doctor and the fault engineer seek to know why a system malfunctions so that they can fix it. This diagnostic use of causal reasoning is one of the most widely studied.

Finally, lawyers seek to know why for yet another purpose. Our profession uses the why question as a means of knowing what occurred. Let me explain how it works. If I know the motives of potential actors, I have a pretty good means of predicting what their actions have been or
will be. It is for this reason that one of the oldest guideposts in the
detection of crime is "who benefits."

So scientists seek broader knowledge of underlying rules, doctors
diagnose malfunctions, and lawyers use causal reasoning as a means of
explaining what has happened or will happen. In this part, I will take a
detour first to discuss briefly these three uses of causal reasoning. Then
I will revert to my customary pattern of first laying out the methodology
and then analyzing that methodology with an eye to explication and
evaluation.

B. Uses of Causal Reasoning

Wesley C. Salmon, in his book on causal reasoning in science, points
out that people want to know why events occur.\textsuperscript{141} Salmon uses the
example of the tides.\textsuperscript{142} Man has known since earliest marine history
that the ebb and flow of the tides correlates with the position and phases
of the moon. The co-variation or frequency of correlation was well
known. But it was not until Newton's causal theory of gravity was
advanced that these correlations were explained. Once mankind had a
gravitational theory, however, it could reason well beyond the
phenomenon of the tides. Rather than predicting solely in areas where
data had already been collected (tide tables and almanacs), scientists could
use the gravitational theory to predict the precise year in which Halley's
comet would return\textsuperscript{143} or to forecast the speed necessary for a rocket to
attain earth orbit for the first time.\textsuperscript{144}

The American philosopher of science Charles S. Peirce, in the early
1900s, developed a paradigm of scientific discovery.\textsuperscript{145} In Peirce's
model, the scientist first develops a causal theory about how some system
under consideration functions. For example, the solar system works as
it does because of the effects of gravity. This process he called
abduction, and the word has recently gained currency denoting the
thought process in which potential explanations are developed.\textsuperscript{146} He

\textsuperscript{141} WESLEY C. SALMON, SCIENTIFIC EXPLANATION AND THE CAUSAL
\textsuperscript{142} Id. at 14.
\textsuperscript{143} Id. at 11-12.
\textsuperscript{144} As I shall discuss infra Section IV.I, possession of a causal theory also helps
avoid mistakes that use of the pure probabilistic model would occasion.
\textsuperscript{145} See NICHOLAS RESCHER, PEIRCE'S PHILOSOPHY OF SCIENCE 2-3, 41-51
(1978).
\textsuperscript{146} See ABDUCTIVE, supra note 59, at 2; EUGENE CHARNIAK & DREW
MCDERMOTT, INTRODUCTION TO ARTIFICIAL INTELLIGENCE 21-22, 453-54 (1985)
[hereinafter CHARNIAK]; HOLLAND ET AL., supra note 66, at 89; Paul Thagard,
believed that once an hypothesis was formed, scientists used deductive methods to determine what events should follow from the theory (e.g., if gravity is at work, then planetary orbits should be more or less circular) and then used inductive methods to test whether those events did in fact follow (e.g., observation indicates that these orbits behave as predicted). The entire sequence he called retroduction. If Peirce is right—as he seems to be in that the hypothetico-deductive method is a recognized scientific approach—then the scientific method is at least initially based on causal reasoning. It is certainly the case, as Salmon points out, that a causal explanation is the objective of scientific search.

In medical and other fault diagnosis situations, practitioners are required to reason causally. If one examines the doctor’s thought process, it goes something like this: “The patient has symptoms including fever, spots, and dehydration; disease X causes fever, spots, and dehydration; no other common disease causes this particular combination of symptoms; therefore, it is likely that the patient has disease X.” Virtually all medical diagnosis is made this way, and it must have some utility or we would not enjoy the relative good health which characterizes this age.

Finally, closer to home, there is compelling evidence that jurors decide law suits by means of causal reasoning. Nancy Pennington and Reid Hastie conducted a series of experiments designed to determine how jurors decide their verdicts. In the first experiment Pennington

---

148. See RESCHER, supra note 145, at 3.
149. SALMON, supra note 141, at 9.
150. See ABDUCTIVE, supra note 59, at 1-9.
151. I hasten to point out that my degree lies in law, not in medicine. Anyone looking for reliable descriptions of medical symptomatology should look elsewhere.
152. See ABSTRACTIVE, supra note 59, at 1-11; CHARNIACK & MCDERMOTT, supra note 146, at 453-65.
153. Medical tests, except those that pinpoint specific disease organisms, are generally simply more refined ways of identifying symptoms. X-ray, ultrasound, cat scan and other tests can actually picture diseased members, but even then the doctor is conjecturing about causal mechanism at work.
presented people drawn from the actual Boston jury pool with a videotaped murder trial and the instructions which a judge would provide at the close of the case. She then asked the jurors to talk aloud as they reasoned their way to a decision. These verbal descriptions of the jurors' thought processes were recorded and transcribed. They were then systematically coded to classify the types of statements made. What emerged were stories depending heavily on judgments about causal motivation. Further, jurors told one of several basic stories, and the story that they told correlated perfectly with the verdict that they reached.\textsuperscript{155} Thus, those whose verdict was that the defendant was guilty of murder in the first degree told a story of the events leading to the killing which was different from that told by those who would have acquitted the defendant. They marshalled different facts and, most importantly for my purposes now, they saw different motives as underlying the fatal fight. Those favoring murder one saw causes consistent with premeditation; those favoring acquittal saw self-defense as the cause for the defendant's actions. I will discuss below the influence of the juror's causal beliefs on the way the evidence was not only classified but even remembered.\textsuperscript{156}

The issue that is important here is that the process in which the legal system asks its decision makers to engage is very like that involved in diagnostic decision making discussed above. The legal system frequently makes state of mind a key issue in the case: negligence, wilfulness, premeditation, carelessness. It almost never provides direct evidence of that state of mind. An admission such as, "It's all my fault; I didn't look where I was going," is very rare evidence indeed. Instead, the system presents the decision maker with symptomatic evidence of the state of mind of the participants in a particular set of events: observations of other people, gaps in the parties' current testimony, etc. The system asks adjudicators—judges or juries—to reconstruct from this symptomatic evidence the underlying mental state that caused the parties to act as they did. The system requires reasoning from symptoms (current evidence) to causes (underlying mental states). Lawyers argue causally to the fact finder. "This evidence only makes sense if you accept that the defendant was not paying sufficient attention to what he was doing."

So there are three important fora in which at least part of the basic mode of reasoning is, in Pierce's phrase, "abductive." Scientists, doctors, judges and jurors reason to some extent in the following mode: "I see the following result before me: patient symptom, event in nature,


\textsuperscript{155} Pennington, supra note 54, at 148, 156.

\textsuperscript{156} See infra Section IV.I.
or legal evidence. As far as I know, only X cause is likely to have produced that result. Therefore, I will hypothesize that X cause did produce that result."

I submit that at least in some instances, practicing lawyers reason in precisely this manner as well. "That will be a good argument to make to Judge X because Judge X tends to favor consumers over merchants in cases like this, and therefore, he will buy this consumer-oriented claim." I think this kind of explanatory thinking frequently characterizes a lawyer's analysis of the likely thought processes of a decision maker. Since those thought processes, unlike their overt results, are not directly accessible, lawyers must reason on these issues from the outward evidence backward to the likely cause. Given an almost certain dearth of direct evidence on how a particular decision maker will decide precisely this issue, causal reasoning is called upon to bridge the gap. Like the fact finder reconstructing the defendant's state of mind from the evidence at trial, the lawyer attempts to construct the likely state of mind of the adjudicator responding to the evidence she has marshalled and the argument she has made.

C. Methodology

Let me provide an example in which the lawyer whose decision making methods I have followed throughout this Article might use causal reasoning to decide a problem. I will suppose that the lawyer is representing a criminal defendant on a charge of simple burglary. In her state, as in most, one of the elements of burglary, which distinguishes it from trespass, is that the alleged perpetrator intended to commit a felony while on the premises. She has marshalled very substantial evidence that her client was severely intoxicated when the police found him on the premises. Since burglary is a "specific intent crime," requiring proof of the intent to commit a felony, the right degree of intoxication is a defense (arguably precluding the formation of the requisite intent). If the alleged perpetrator was so drunk that he did not know where he was or if he was so drunk that he was simply looking for a place to sleep, he is not guilty of burglary.

The decision that the lawyer must make is whether she should recommend that her client insist on his right to jury trial or waive a jury trial to be tried by the judge alone.157 I will assume as is the case in most jurisdictions that the lawyer knows to which judge the case has been

157. I will assume that a plea bargain is not available or is undesirable for some reason. Also, please note here that the ultimate decision must rest with the client because of the rights implicated.
assigned before this decision must be made irrevocably. I will also assume that she has sufficient experience to know the general characteristics of venires in her area so she has a good idea of the mix of jurors she is likely to draw. The question is, in this particular case, which fact finder is more likely to favor her client. There is some evidence that lawyers may be pretty good at making these decisions. Kalven and Zeisel, in the American Jury,\(^\text{158}\) report that in criminal cases in which the lawyer elected to retain the jury, the judge’s hypothetical verdict tended in a significant measure to be harsher than that rendered by the jury.\(^\text{159}\)

Let me theorize about how such a decision might be made using causal reasoning. By contrasting the set of facts that might cause the judge to accept her intoxication defense with the set that might cause the jury to do so, the lawyer attempts to select the most plausible scenario for acquittal using this defense. The approach that she should use can be summarized as follows. First, as with induction, she needs to identify phenomena that co-vary with case outcomes. In her past experience, what features of cases, such as strength of legal arguments, client credibility, weakness of eyewitness evidence, etc., characterized those that led to acquittal? She is looking for specific forms of co-variation that identify possible causal factors.

Having identified possible causes, she will have to rely on three tests to determine whether the candidates are merely co-occurring events or whether they remain in the running as possible causal explanations. One of these tests depends on contextual knowledge akin to that which figured also in induction. Do the proposed causes mesh with her understanding of how the courts work? The second test is easier to apply though surprisingly important. She needs to be sure that the proposed cause precedes the result in time. Finally, she must determine whether some

\(^{158}\) HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY (1966) [hereinafter JURY]. In this widely recognized empirical study of jury behavior, Kalven and Zeisel sought to measure the differences between the jury and the presiding judge as “judges of the case.” To do this they asked presiding judges in jury trials to record, before the jury returned, the verdict that they would have reached had they heard the matter without a jury.

\(^{159}\) With regard to their study, Kalven and Zeisel noted “[t]he jury is less lenient than the judge in 3 percent of the cases and more lenient than the judge in 19 percent of the cases.” \textit{Id.} at 58-59. Of course these data give us only half the story. They did not study whether advocates displayed similar predictive powers when those advocates elected to waive jury. Such a study would be very difficult to conduct because it would involve the empaneling of shadow juries in situations where the defense had waived. But without these data, it is impossible to know where predictions are really good or whether juries always favor defendants more heavily than do judges. This is an example of the four cell co-variation problem. \textit{See supra} note 136.
link exists between the putative cause and the result she seeks to predict. For example, if the fact finder could not have learned of the evidence which is assumed to be the influential cause, then that theory of causation must he dismissed.

If the lawyer is convinced that she has identified a potential causal relationship, she must determine whether that relationship is the true cause or best explanation of the result. She will see whether the relationship explains or "covers" all the facts at her disposal. If some key phenomenon, for example a relationship between client demeanor and verdict, is not explained by the theory, then that theory is suspect. Finally, she will see whether this explanation is the most economical or simplest available. As with previous methodologies, let me examine these steps one at a time.

D. Mill's Co-Variation Tests

John Stuart Mill provided a model of causal reasoning in 1852.\textsuperscript{160} The version of his model that I will discuss is that set forth in Brian Skyrms's excellent introduction to decision modeling, *Choice and Chance*.\textsuperscript{161} The approach here is one of systematic co-variation observation.\textsuperscript{162} Only by noting co-variation can the lawyer lay the foundation for identifying cause. Mill's tests are designed to identify two separate kinds of causality. A "necessary cause"\textsuperscript{163} is one which cannot be absent when the result caused is present. For example, oxygen is a "necessary cause" of combustion. By contrast, a "sufficient cause" is one which cannot be present when the result caused is missing. Oxygen is not a "sufficient cause" of fire, thank goodness, because oxygen is present in many cases in which fire is absent. In the combustion example, one

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{160} See John S. Mill, A System of Logic (8th ed., Harper & Bros. 1874).
\item \textsuperscript{161} Brian Skyrms, Choice & Chance: An Introduction to Inductive Logic 75-111 (3d ed. 1986). Skyrms credits much of his analysis to Georg H. von Wright, A Treatise on Induction and Probability (1960). For another discussion of Mill's approach, see Cohen, supra note 51, at 144-51.
\item \textsuperscript{162} See supra Section III.G. As pointed out there, human beings are not generally very good at co-variance detection.
\item \textsuperscript{163} I will use the term "cause" throughout this discussion of Mill's methods. This is a little unfair to both Mill and Skyrms. Mill did not claim that his method isolated "causes." Rather he said that it identified, "effect[s], or cause[s], or necessary part[s] of causes." See Mill, supra note 160, at 225. See also id. at 222, 224, 229, 233. Skyrms substitutes the term "condition" for "cause" throughout his text and thus fineses the issue of causality, although he points out that Mill thought he was testing for causes. Skyrms, supra note 161, at 88. Therefore, when I discuss the fact that co-variance cannot insure causality, infra Section IV.E, I am to some extent battering a straw man of my own creation.
\end{itemize}
\end{footnotesize}
"sufficient cause" for fire would be the presence of the right combination of fuel, oxygen, and a match. Therefore, the result of a causal process may be present when a "necessary cause" is there, but if a "sufficient cause" is present the result must occur.

Once this difference between the two types of causes is recognized, it is possible to reason backward using a presence/absence table to see which possible causes that co-occur with a result may actually be a "necessary" or "sufficient cause." I will stick with the fire example first. If you will look at Table I, you will see such a table composed after the occurrence of three fires. Each fire is shown in rows marked occurrence #1-#3. Each possible causal agent is listed across the top and is assigned a column. The result, a fire, is listed in the last column.

**Table I**

<table>
<thead>
<tr>
<th></th>
<th>Gas</th>
<th>Wood</th>
<th>Oxygen</th>
<th>Match</th>
<th>Water</th>
<th>Sparky</th>
<th>Fire</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>#2</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>#3</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

At the first fire, gas was present (P), wood absent (A), oxygen present (P), match present (P), water present (P), and Sparky, the fire house Dalmatian, present (P). The fact that fire resulted is shown in the last entry, fire present (P). Based on these data all of the possible causal agents with the exception of wood is a candidate for a "necessary cause." They were present when the fire eventuated. However, in occurrence #2, water is missing (A) while the fire is still present (P). From this information, it is safe to conclude that water is not "necessary" to a fire. Fires can occur without water being present. The system seems to be working well, but now examine occurrence #3. Here suddenly gas which

164. This type of table and particularly the values assigned to disjunctive (or) and conjunctive (and) combinations are modeled on the "truth tables" integral to Boolean logic. For a discussion of this relationship, see SKYRMS, supra note 161, at 75-82. Boolean logic is named for George Boole, an English mathematician who lived from 1815 to 1864. Most lawyers are familiar with his symbolic algebra of connections, "not," "and," and "or," but not necessarily with its source. It is this logic that underlies searches in full-text databases such as "Lexis" and "Westlaw." For an additional discussion of this form of logic, see RUDY RECKER, MINDTOOLS: THE FIVE LEVELS OF MATHEMATICAL REALITY 207-12 (1987).
had been present at both previous fires is now absent (A). Equally surprising is wood, which had been absent at the previous fires and is now present (P). This serves to illustrate that "necessary causes" may be disjunctively compound. That is, the table needs a new column, see Table II, which is titled (Gas or Wood). With this column, the observations previously made are reconciled with this new information. So, after these three observations, you can see that the possible necessary causes have been whittled down to (Gas or Wood), Oxygen, Match, and Sparky, the firehouse Dalmatian. Your instincts may tell you that Sparky ought to be eliminated as a possible cause or perhaps should never have made the table in the first place, but I want to reserve until later a discussion of why this may be so. For the moment, Sparky stays put.

Table II

<table>
<thead>
<tr>
<th></th>
<th>Gas or Wood</th>
<th>Oxygen</th>
<th>Match</th>
<th>Water</th>
<th>Sparky</th>
<th>Fire</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>#2</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>#3</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>A</td>
<td>P.</td>
<td>P</td>
</tr>
</tbody>
</table>

Now that there is a list of possible "necessary causes," I need a method to test for any "sufficient causes." You will remember that if a "sufficient" cause is present, the result must occur. Therefore, the obvious test is to look for instances in which some or all of the possible "necessary causes" are present, but fire does not occur. Look to Table III. I have listed one column for (Gas or Wood), one for Oxygen, one for the Match, and one for Sparky. In the first occurrence, when there is a fire, all four possible "necessary" and "sufficient causes" are present. In the next instance (row #2), in which there is no fire (A) all of the possible causes are present except Oxygen. This reveals that (Gas or Wood), the Match and Sparky cannot be "sufficient causes" separately or in combination. They occurred, and fire did not. Finally in occurrence #3, I test whether Oxygen is a "sufficient cause." Removing all the other possible causes, you see that Oxygen by itself does not cause the fire (A).

165. Knowing what I do about fires, I might be tempted to short circuit the process and just call it generically carbon based fuel, but remember, at this state I know nothing about fire beyond the three examples shown.
The question is what conclusion should I draw now. Two are available. The first is that none of the possible causes is a "sufficient cause." This would be a tenable proposition. There are many occasions in which some circumstances are the "but for" or "necessary causes" of an event, but require other events to intervene before the result occurs. A banana peel in front of the produce section is the "necessary cause" of my fall on a banana peel. (I would not have fallen if I had not stepped on the peel.) But it only becomes part of the "sufficient cause" when coupled with my inattention to where I am putting my feet. I could also hypothesize, however, that some combination of the possible causes already identified constitutes a "sufficient cause." See Table IV. If I add a column to the table for the conjunction of ((Gas or Wood) and Oxygen and Match and Sparky), you will see that it correlates perfectly with the final column for fire. Thus, you and I could conclude that "All of the Above" are the "sufficient cause" for fire.

### Table III

<table>
<thead>
<tr>
<th></th>
<th>Gas or Wood</th>
<th>Oxygen</th>
<th>Match</th>
<th>Sparky</th>
<th>Fire</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>#2</td>
<td>P</td>
<td>A</td>
<td>P</td>
<td>P</td>
<td>A</td>
</tr>
<tr>
<td>#3</td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Mill's tests go on to more complicated patterns of data and combinations of possible causes, but this discussion has been sufficient (and necessary?) for my purposes. Let me turn now to how the insights here will assist the lawyer in deciding whether to waive jury in her
client's burglary case. I will assume that in the past she has tried eight cases before this judge involving an intoxication defense to a burglary charge. In four she waived jury and in four she kept the jury. I will further assume that those juries were very like the venire she faces now. See Table V for a summary of her experience. The columns labeled burglary and intoxication are self-explanatory and are true in each case. The column labeled Jury indicates when she waived (A) and when she retained (P). The column entitled Legal Claim indicates those instances in which she was able to make a very solid legal argument that her client fell within the defense (P) and when she could not make such an argument vigorously (A). The Sympathetic Defendant column indicates those instances in which her client was appealing (P) and those in which the client was not (A). The result, acquittal column shows a (P) when the client was acquitted and an (A) when he was not. Of course what my example shows is the very unlikely occurrence of a perfect co-variation.

Concentrate on the last three causes: Jury, Legal Claim, and Sympathetic Defendant. In the two instances (cases #1 & #2) where jury was waived and a strong legal argument was made, an acquittal was obtained. However, in case #6 where the jury was retained, and a strong legal argument was made, but the client was unsympathetic, acquittal did not occur. In terms of Mill's tests, a strong legal case by itself is not a "sufficient cause" of acquittal. However, the conjunction of a strong case and jury waiver may be a "sufficient cause" of such an acquittal. Also, a strong legal case cannot be the "necessary cause" of acquittal because an acquittal took place in case #7 when such a strong case was missing. On the other hand, it would appear that the disjunction, either a strong legal case or a sympathetic defendant, is "necessary" to acquittal because in the two cases (case #4 and case #8) where both were missing an acquittal was never obtained.

Thus, based on Mill's tests, the lawyer would form the following tentative causal hypothesis: "If I am to win this case, I must have either a strong legal theory or a sympathetic defendant. If my legal argument

166. This is highly unlikely, even in one judge jurisdictions, given the frequency of plea bargains and the relative infrequency of this defense. However, the assumption will illustrate my point, so I will make it for didactic purposes.

167. It is unlikely in two respects. First, the correlation is too good to reflect reality. Second, I do not believe that the legal system functions as I have depicted it. I think judges respond to sympathetic defendants with about the same frequency that juries do, and I think that a compelling legal claim frequently has a compelling factual foundation and is thus as persuasive to a jury as to a judge. But again for didactic purposes this model is helpful, and one closer to reality would require too much explanation of my underlying prejudices about jury waiver choice, which is not the topic here.
is strong, but my defendant unsympathetic, I should opt for the judge. If my defendant is sympathetic, but my legal argument weak, I should choose the jury. If both are strong, it does not matter which forum I elect."

These conclusions are almost certainly too simplistic, but they do serve to illustrate that co-variation data can serve both to eliminate a causal hypothesis—here a sympathetic defendant does not always win—and to home in on hypotheses which may be worth further testing. It is important to keep in mind that, regardless of Mill’s terminology, the fact that data do co-vary is no true guarantee of causation. (I discuss this issue below.)

I will stop here to introduce the important caveat that there is very good reason to believe that decision makers do not in fact perform detailed co-variation analysis of the kind called for in Mill’s tests when they develop causal theories. As I have discussed above, people in general are bad at this kind of co-variation analysis and quite adept

---

168. The lawyer here might add a further explanatory layer and assert that the judge is more amenable to legal persuasion than the jury because of the judge’s greater familiarity with legal arguments or for some other reason. See ROGER C. SCHANK, EXPLANATION PATTERNS: UNDERSTANDING MECHANICALLY AND CREATIVELY 14-16 (1986) (distinguishing between explanations that make sense of the world (explain reasoning processes) from those that advance a theory about why the world works as it does).

169. See supra Section III.G.
One explanation of these facts currently enjoying considerable respect is that people do notice surprising failures of co-variation expectations. They learn from those failures of prediction but do not store the information developed as rules of the kind produced here (fuel + oxygen + ignition = fire). Rather they store the learning as stories (when I lit that fire) of which subsequent events remind them and from which they draw lessons for future conduct. So the lawyer in my example may be overly observant in the pattern recognition capacities with which I have credited her.

E. Separating Co-Variance from Causality by Contextual Knowledge

Whether the lawyer does well or poorly at noticing co-variations, all she has done to this point is to identify some possible causes of the result she seeks. If causal theories do contribute to the understanding of uncertain situations, then it is clear that mere co-variance will not insure that causal links have been identified. Yun Peng and James A. Reggia in Abductive Inference Models for Diagnostic Problem-Solving make this point through use of an area diagram. See Figure 13. The upper left hand quarter contains the instances in which both of the co-varying phenomena are present. Within that group of phenomena which may be seen to co-vary, only some (those illustrated by the circle) will do so because one event causes the other. Part of co-variation will almost certainly be purely coincidental and part may reflect a third, as yet unknown, cause. Other tests, beyond co-variation must be imposed to determine whether the lawyer has identified possible causal agents.

First, it is the context of the discipline within which the causal connection is sought that tells the decision maker whether her proposed cause is tenable. Let me discuss how this meshes with the two types of "cause" identified in Mill's test above. It is relatively easy to see that a possible "necessary cause" might not be a cause at all. You will recall

170. See Amos Tversky and Daniel Kahneman, Causal Schemas in Judgments under Uncertainty [hereinafter Causal Schemas], in HEURISTICS, supra note 27, at 117-18. See also HUMAN INFERENACE, supra note 27, at 10, 113-38.
172. See ADBDUCTIVE, supra note 59, at 101-04; SALMON, supra note 141, at 121. See also ELLERY EELLS, PROBABILISTIC CAUSALITY 56-57 (1991) [hereinafter EELLS: PROBABILISTIC] ("It is famous that 'correlation is no proof of causation' and it is also true that causation does imply correlation.").
173. See infra Section IV.J.
that the definition of a "necessary cause" is a condition which cannot be absent when the result is present. My example was oxygen ("necessary cause") which cannot be absent if combustion (result) is to take place. You will also recall, however, that a "necessary cause" may be present when the result does not occur. Oxygen is present in many situations in which there is no fire. So stripped of the causative language and looked at purely as a matter of co-variation, one phenomenon (oxygen) may be present (vary) whether or not another phenomenon (combustion) occurs. Any number of phenomena answer this description which I certainly would be unlikely to label causal. The world’s oceans for example are present whether or not fire occurs.\textsuperscript{174} Fire never takes place, as far as I know, without the existence of the those oceans. But I would not hold that oceans cause fire. What makes me believe that oxygen is causative of fire is only to a small extent its co-variation pattern. That pattern makes oxygen a candidate for the status of "necessary causal" agent, but does no more. It is the scientific context of physics and chemistry that prompts me to label oxygen truly a "but for" cause of fire.

\textsuperscript{174} Although not necessarily in their present form, however, if we are to believe theories of continental drift and plate tectonics.
The lawyer has concluded that a strong legal argument or a sympathetic defendant are the necessary disjunctive "cause" of an acquittal. Our understanding of the legal context does not rule out this causal hypothesis. We expect both legal merit and community values to influence legal decision making.

It is a little harder to see that the role of Mill's "sufficient cause" is also only to propose candidates for consideration as causal theories. After all, you will recall that to constitute a "sufficient cause" a phenomenon cannot be present when the result is absent. This definition does not insure that there is perfect co-variance between the proposed agent and result. It is possible that the result could be present without the putative cause. However, it does insure that whenever the proposed cause is present, the result occurs. Again the example from the combustion discussion comes to mind. Doesn't the fact that fuel, oxygen, and a match are always present at a fire guarantee that these are the causes of that fire? Unfortunately no. What guarantees the fire is the physics and chemistry of combustion, not the presence of co-variation.

My original combustion example contained a co-variance whose causal import seems dubious. It was the presence of Sparky, the firehouse Dalmatian. Sparky's presence correlated perfectly with the result and obeyed all the other rules for a part of a "sufficient conjunctive cause." But leaving aside the possibility that the fire chief may be both an animal fancier and a pyromaniac, in which case we would have unearthed an explanation for Sparky's presence, I would set aside this correlation because I have no adequate causal theory to explain it. It does not fit in the context of combustion.

Are the lawyer's "sufficient causal" hypotheses contrary to common understanding of the trial court context? I don't believe so. It is not inherently unreasonable that a judge trained in law would be more persuaded by legal arguments, while jurors brought in to reflect the community should be more amenable to appeals to sympathy.

175. All that is guaranteed by status as a "sufficient cause" is that the first cell in the table (Present/Present) and the second cell (Present/Absent) will always correlate perfectly. That is when the "cause" is present, the "effect" will be present, and when the "cause" is present, the "effect" cannot be absent. It tells nothing about the third and fourth cells. It tells nothing about what happens when the "cause" is absent. The "effect" may occur anyway. For example, every time I have looked out the window today, the sun has been shining. My gaze out the window meets the test as a "sufficient cause" for the sunshine. But for all I know, the sun has elected to shine all day today, with or without my assistance.  
176. It is, of course, possible that some such theory would emerge in the future. See my discussion of context in the inductive part of this Article, supra Section III.D.
F. Causality Implies Precedence

Once causal candidates make sense in the context of the court system, the lawyer must be certain that the phenomenon which she views as causal precedes the phenomenon which she sees as a result. This point seems self-evident but involves two issues which need attention: one theoretical, one practical. Probability theory, which is concerned only with correlations, not causes, has no such time direction requirement. So the fact that two phenomena are correlated is a symmetrical relationship, whereas the fact that one causes the other is not. This is not the only reason why conventional probability theory cannot adequately describe causality (see my discussion below), but it is an important disability.

There is evidence that decision makers assume temporal order far too quickly. Amos Tversky and Daniel Kahneman posed the following problem to experimental subjects:

Let $A$ be the event that before the end of next year, Peter will have installed a burglar alarm system in his home. Let $B$ denote the event that Peter's home will be burglarized before the end of next year. Let $\overline{A}$ and $\overline{B}$ denote the negations of $A$ and $B$, respectively.

Question: Which of the two conditional probabilities, $P(A \mid B)$ or $P(A \mid \overline{B})$, is higher?

Question: Which of the two conditional probabilities, $P(B \mid A)$ or $P(B \mid \overline{A})$, is higher?

The researchers report their result: "A large majority of the subjects (132 of 161) stated that $P(A \mid B) > P(A \mid \overline{B})$ and that $P(B \mid A) < P(B \mid \overline{A})$, contrary to the laws of probability." The mistake the subjects made, of course, is that they neglected to take the direction of time into account in formulating their answers. A burglary only makes installing a burglar alarm more likely if the burglary precedes the installation. But conversely, the installation of a burglar alarm only makes a burglary less likely if it precedes the burglary. It is the failure to have a consistent

177. See Eells: Probabilistic, supra note 172, at 239-77.
178. Id. at 239-44.
179. See Causal Schemas, supra note 170, at 123. Problems of temporal ordering of causal inferences are frequent enough to have earned a name and are called Turoff's problems. If you need a reminder on the meaning of the notation I am using, see supra note 14.
hypothesis about which came first that led the respondents into error here. Respondents neglected caution about temporal priority.

In the lawyer's example, temporal priority is not a problem. The decision to waive or retain the jury coupled with either making the strong legal arguments or demonstrating the clients' sympathetic characters all preceded trial outcomes in her previous cases.

G. Cause and Effect Must Be Linked

Finally there must be an actual causal link of some type between the putative cause and its effect. The precise nature of the causal link is itself context sensitive to a degree. In the physical sciences, a cause probably is best defined, as Salmon does, as that which brings about a change in a physical process. Thus a fire might heat and thus cause a change in a liquid, or one mass might influence (by striking, gravity, or some other means) or cause a change in the path of a second mass.\textsuperscript{180} In understanding human reactions, notions of goals and motives figure strongly in causal schemes, as Pennington and Reid point out. So an actor is said to be influenced—i.e., to have his behavior caused—by a combination of these internal goals and the events which befall him.\textsuperscript{181} Attempting to work out a general causal schema, Roger Schank and Robert Abelson identified five rules which they believed described a complete causal syntax:

\begin{itemize}
  \item CS1 Actions can result in state changes.
  \item CS2 States can enable actions.
  \item CS3 States can disable actions.
  \item CS4 States (or acts) can initiate mental states.
  \item CS5 Mental states can be reasons for actions.\textsuperscript{182}
\end{itemize}

Inherent in all those ideas is interplay or interaction between the possible cause and the result. If the lawyer had provided the jury with no way of learning of her client's sympathetic nature, then she would be ill-advised to assume that that nature influenced the jury to acquit.

Assuming she has not made this error, she now has causal hypotheses that accord with her prior experience, are congruent with her understanding of the court room context, are in the right temporal order,

\textsuperscript{180} See Salmon, supra note 141, at 139-47. Salmon also identifies another sort of cause type which he calls propagation. For his discussion, see id. at 147-57.

\textsuperscript{181} Pennington & Hastie: Decision, supra note 154, at 5; Pennington & Hastie: Evidence Evaluation, supra note 154, at 243-44.

and demonstrate some link between the cause and effect. But, unfortunately, an hypothesis could meet all those tests and yet not be the true explanation of the result under study. Look now at Table VI. I have added a new column entitled “Female.” Let me assume that no acquitted defendants were female (A) and all convicted defendants were (P). The lawyer now has a rival, and at least as powerful, causal hypothesis. It appears that judges and juries in her locale may, in burglary intoxication cases, discriminate against women and in favor of men. How is she to choose between her prior hypothesis (which I will call the “case” hypothesis) and this (“sex based”) one? The next three sections address this issue of selecting the most plausible explanation from among competing hypotheses.

**H. Assessing Causal Plausibility**

Philosophers and programmers working in the realm of artificial intelligence have recently made major contributions to this selection process. Yun Peng and James A. Reggia, in *Abductive Inference Models for Diagnostic Problem-Solving*, and Thagard, in *Explanatory...

---

183. I will assume, without repeating the prior analysis, that the sex-based theory meets the tests of context, temporal priority, and linkage just discussed.

184. It may be that a social stigma attached to a female drunk, while male drunkenness is more tolerantly viewed. Women then would be punished for attempting to use intoxication as a defense.

185. *ABDUCTIVE, supra* note 59.
Coherence,\textsuperscript{186} have within the last two years successfully modeled decision processes which took causal theories as inputs and rated their plausibility. Both treat causal relations and explanations as "primitives." (A primitive is a given in the computer's logic structure. It must be supplied by the programmer and cannot be computed from data fed to the program. Because of this feature, a primitive need not be defined.) They do not say how the explanations are developed or causes hypothesized. Rather they take the hypothetical causes or explanations as givens and seek to rank them. They step into the process at precisely the point our hypothetical lawyer has now reached. The tests that they elected to apply share some important similarities that will help the lawyer here.\textsuperscript{187} Peng, Reggia, and Thagard all suggest that she submit her hypotheses or causal explanations to the following two additional tests to determine which is the best available explanation of her data. Do these theories "cover" all the relevant data that she has observed?\textsuperscript{188} And do the theories offer the simplest explanation of those data available?\textsuperscript{189} Both criteria are data driven; that is, they test theories against the actual observations that the lawyer has made.\textsuperscript{190} I will take each separately.

1. COVER THEORY

If you will refer back to Table VI, you will see that each causal hypothesis that the lawyer advances covers all her data. There is no case here in which a weak legal case won the heart of her judge or where a

\textsuperscript{186}Thagard, \textit{supra} note 146.

\textsuperscript{187}This is interesting because the causal theories that they are appraising lie in very different realms. As the name of their book discloses, Peng and Reggia are interested in diagnostic applications of causal reasoning. They seek to model medical diagnosis in the program they describe. Thagard, on the other hand, is interested in ranking rival scientific theories or forensic arguments. Nevertheless, both share the task of finding the best causal hypothesis.

There is an additional methodological point of interest. The programs that embody their theories are all written as parallel nets. This represents admissions in both cases that the reasoning required to make these assessments must keep multiple variables operative at once, that the interaction of these variables must be parallel, and that large numbers of computations are required. They are modeling highly complex and non-linear mental processes.

\textsuperscript{188}See \textit{ABDUCTIVE}, \textit{supra} note 59, at 7, 22-24, 255-58; Thagard, \textit{supra} note 146, at 436-38.

\textsuperscript{189}See \textit{ABDUCTIVE}, \textit{supra} note 59, at 7, 22-24, 255-58; Thagard, \textit{supra} note 146, at 436-38.

\textsuperscript{190}ABDUCTIVE, \textit{supra} note 59, at 28-33 (answers to the questions generated by the system used in an hypothesis and test model); Thagard, \textit{supra} note 146, at 437-38, 442-43 ("evidence" provided by the programmer is the sole initial energy source for the system).
jury was not stirred by sympathy for an appropriate defendant. Nor is there any case in which a man was convicted or a woman acquitted. But what if on her next matter, the case at issue here, the client is female and in the lawyer's judgment sympathetic, she does not waive jury, and yet the defendant is convicted? See Table VII for this new situation. How should the lawyer deal with this aberration? First, of course, she should reappraise her sense of the sympathetic nature of her client. She may just have a problem of misclassification. But assuming that she remains convinced that she classified correctly, Peng and Reggia would give her different advice from Thagard. Thagard would simply decrease the strength of the “case” theory she advanced, but if the better “sex based” theory had not emerged, he would not exclude this “case” hypothesis.\(^{(191)}\) Peng and Reggia would regard the “case” hypothesis by itself as untenable whether or not the “sex based” theory was available. Some other hypothesis would have to be coupled with it to explain the otherwise unexplained data.\(^{(192)}\)

\(^{(191)}\) Thagard, \textit{supra} note 146, at 438. At some point acceptability might be so reduced that the theory would not pass muster, but the mere existence of unexplained evidence is not immediately fatal.

\(^{(192)}\) \textit{Abductive}, \textit{supra} note 59, at 7, 104. Their program assumes that no symptom can occur which is not explained by a disorder. They would permit her to salvage the hypothesis by joining it to another that explained the anomaly, but by itself it is discredited.
This difference in approach may well be traceable to the different systems that the theorists seek to model. Peng and Reggia are modeling diagnostic systems. They assume by definition that all symptoms are relevant and are explained by causes already identified in the system. The price of this assumption is a necessarily relaxed view of simplicity as I will discuss below. Thagard is modeling scientific and forensic argument. He seeks to find the best theory among competing contenders, and he is working in a universe where each piece of evidence cannot necessarily be presumed germane. Thus he has a broader tolerance for anomalous conditions.

In any event, both agree that the lawyer ought to look with a jaundiced eye at her "case" hypothesis should it fail as described. Perhaps she needs to add another feature, such as the strength of the prosecution's evidence on intent, if she seeks to improve that causal model. A failure to cover known data should trigger hypothesis reappraisal, if not certain rejection. Here the "sex based" hypothesis would clearly be favored on these data. It provides the best cover.

2. SIMPLICITY

For purposes of the next test, let me revert to the data set forth in Table VI. This defendant has not been tried and the data give no basis to distinguish between the theories on the basis of cover. The notion of simplicity as a desirable feature of explanatory theories dates at least from William of Occam, an English philosopher in the 1200s. "Occam's Razor" though honed on the strop of theological dispute, has come to mean generally that "the fewest possible assumptions are to be made in explaining a thing." Several arguments justified application of the

193. Id. at 5-6, 133. If an available explanation does not account for all symptoms, something else must be amiss.

Finally, [the mandatory causation assumption] . . . is a very basic assumption in a diagnostic world. It combines the fact that any manifestation cannot be present unless it is caused by some disorder and the closed-world assumption that all possible causative disorders are explicitly listed in D. To the authors' knowledge, this cause-effect relationship assumption has not been made in the past.

Id. at 106.

194. He tests his program on scientific controversies (Darwinian evolution versus creationism, for example) and to a lesser extent on evidence from trials. Thagard, supra note 146, at 443-53.

195. For a discussion of the similar handling of anomalies in inductive reasoning, see supra Section III.F.


197. OXFORD, supra note 64, at 819.
simplicity criterion. Peng and Reggia suggest that it is used in deference
to the short term memory limits of humans which preclude more complex
explanatory chains or as a generally successful rule of thumb.\textsuperscript{198}
Thagard and others see it as a measure of coherence, providing inherent
plausibility to the explanation being evaluated.\textsuperscript{199} Peng, Reggia, and
Thagard share a commitment to simplicity although they define it
differently. For Thagard the number of theoretical propositions required
to contribute to the explanation of a result is the inverse measure of
coherence and simplicity.\textsuperscript{200} Peng and Reggia hold a broader view
because, in their system, more than one fault may be required to explain
all the symptoms present.\textsuperscript{201} A simple count of operative causes will
not do. They prefer a test that measures simplicity, or parsimony as they
call it, by the test of irredundance. A parsimonious explanation is one
that involves no more causes than are required to account for the
symptoms present.\textsuperscript{202}

Under either of these approaches, the lawyer should look to see
which of her causal hypotheses was the simplest. On the evidence
available, a conjunct theory which takes into account both the fact finder
and nature of the case or of the defendant requires two explanatory
propositions to explain the results observed while the "sex based"
hypothesis requires only one. Under either test, the "sex based"
hypothesis is the simplest. Even if the "case" theory covered all the
cases, she should nevertheless prefer the explanation that offers the
greater simplicity.

It is evident, of course, that neither of these tests provide the lawyer
an absolute guarantee that the "sex based" hypothesis, with its assumption
about judge and jury prejudice, really reflects the motivation of those
adjudicators. The hypothesis is always subject to disconfirmation through
a failure to cover the acquittal of a subsequently tried woman, or it may
be challenged by another equally simple proposition. Until these events
occur, however, the lawyer would be using good judgment to advise her
female clients that they were likely to fare poorly with this defense.

\textsuperscript{198.} ABDUCTIVE, supra note 59, at 7.
\textsuperscript{199.} Thagard, supra note 146, at 437. For Kenneth S. Friedman, simplicity
means testability. If an explanation may be confirmed or disconfirmed by the imposition
of fewer tests than its competitor, that explanation is simpler and more plausible.
\textsuperscript{200.} Thagard, supra note 146, at 442.
\textsuperscript{201.} ABDUCTIVE, supra note 59, at 262.
\textsuperscript{202.} Id. at 117-20.
I. Should Causal Explanations Be Preferred to Probability Calculations?

Assuming that a causal or explanatory hypothesis meets these tests of co-variance and causal connection, and is optimized using cover and simplicity, would not all decision theorists agree that that hypothesis provides the best basis for prediction in the face of uncertainty? Should not the lawyer rely on the implications of such a tested hypothesis as opposed to those devised from standard probability analysis? After all, if I know what makes an event happen, are not my predictions much better than if I only know its frequency? Surprisingly a substantial group of cognitive psychologists and decision theorists say no.

Included in their ranks are Daniel Kahneman, Amos Tversky, Richard Nisbett, and Lee Ross, very prominent thinkers on, and careful empirical students of, human decision making. These critics offer three basic objections to the causal approach to decision making. These objections are: decision makers' sense of co-variation is too unreliable to permit formation of good causal theories; people prize causal explanations too highly with the result that they will adopt the first plausible argument that comes along; and finally, once a causal theory has been adopted, it will be maintained by decision makers even in the face of substantial refuting evidence.

I have discussed co-variation problems before. There is substantial evidence that people do not do this work well.203 Experiments in the skill of causal attribution make this point. It is possible, simply by changing a decision maker's line of sight in observing a situation, to alter that observer's sense of who caused an event.204 In one experiment, observers watched a conversation. They could see one party well, but the other poorly. They uniformly rated the party they could see well as causing the conversation to proceed as it did. This notion of control through line of sight is, of course, one which trial advocacy teachers help students to exploit in presenting a direct or cross, so this finding should come as no surprise to lawyers. In the case of direct, students are advised to fade into the woodwork behind the jury so that the witness is seen as powerful and credible. On cross, for converse reasons, the students are advised to draw the attention of the jury to themselves. If trial advocacy teachers did not believe that the jury's sense of co-variation could be influenced by these maneuvers, this advice would make little sense. But if the ability to observe causal co-variation is amenable to

203. See supra Section III.G.
204. See HUMAN INFERENCE, supra note 27, at 125.
such subversion, it does give pause to the assumption that Mill’s tests or other co-variation detection devices will prove workable in the real world.

On the other hand, if people are not good at co-variation detection in the causal context, that defect extends to the inductive and decision theory models as well. This criticism does not seem uniquely to disable causal reasoning, but rather to infect any process that attempts to bring past experience to bear on present uncertainty. People may not be very good at doing this, but they have no choice.

The second criticism, however, is aimed specifically at the causal model. You will recall that when I was discussing temporal priority, I cited an experiment showing that subjects favor causal explanations even in the face of temporal ambiguity. A similar experiment shows that this preference persists when a fact could equally be either causal or diagnostic. Tversky and Kahneman administered this problem in the 1970s:

**Problem 9:** Which of the following two probabilities is higher?

- $P(R | H)$ The probability that there will be rationing of fuel for individual consumers in the US during the 1990s, if you assume that a marked increase in the use of solar energy for home heating will occur in the 1980s.
- $P(R | H)$ The probability that there will be rationing of fuel for individual consumers in the US during the 1990s, if you assume that no marked increase in the use of solar energy for home heating will occur during the 1980s.

Sixty-eight of their eighty-three respondents said that rationing was less likely if solar heating was increased than if it was not. Subjects disproportionately saw the failure to save heating oil in the 1980s as making a shortage in the 1990s more likely. But the facts could equally have been read to indicate a mounting energy crisis spanning the two decades in which both increase in solar heating and adoption of rationing were diagnostic symptoms of that crisis. I shall discuss below the implications of the fact that co-varying events can easily be joint symptoms of a third cause rather than having any causal relation to each other. In the meantime, Tversky and Kahneman have shown that the

205. *See supra* Section III.G.
206. *See supra* Section IV.F.
208. Indeed, Tversky and Kahneman suggest that the diagnostic rather than the causal reading is to be preferred because it is unlikely that sufficient savings could be made from solar heating to stave off a true energy crisis.
209. *See infra* Section IV.J.
tendency to assume a surface causal relation is pronounced. People do in fact see causes where there may not be any.

This becomes a particular problem when it is confounded by the third observation that once causal explanations are formed, they are very difficult to shake. I have already discussed this problem as well in the context of induction. It is the issue of accommodating anomalies. You will recall that Holland et al. hypothesized that decision makers retain disconfirmed rules longer than perhaps they should by developing particular case exceptions. There is evidence that those decision makers adjust to unwelcome data in other ways as well. Nancy Pennington and Reid Hastie have shown that mock jurors remember different facts depending upon the stories which they construct to decide their verdicts. Not only do they forget facts that are incongruent with their stories, they remember facts that were not there. In a study of the effects of disconfirming data, Lee Ross and Craig Anderson report that subjects who felt strongly about capital punishment were asked to read two separate "studies" one purporting to provide empirical support for their position and the other supporting the opposing view. The "study" that did not support the subjects' prior positions was attacked by them as methodologically flawed, and the subjects ended the experiment by holding their previous positions more strongly than they had prior to exposure to the "studies."

Such evidence certainly shows that there is a tight link between the explanations that decision makers adopt and the evidence that they elect to recognize. That very different views of underlying facts can give rise to plausible arguments both about what really happened in the past and what should happen in the future comes as no news bulletin to lawyers. That, after all, is why we insist on two plausible theorizers contesting each other's premises and conclusions in our trial courtrooms. On the other hand, this critique is not unique to causal thinking. Resistance to abandoning preconceptions has been demonstrated in many contexts in which causal hypothesizing is not at issue. Tversky and Kahneman report an experiment in which subjects were asked to estimate the percentage of

---

210. See supra Section III.F.
211. PENNINGTON, supra note 54, at 105, 107, 114. See supra text accompanying note 158 for a summary of the research. Jurors holding the defendant guilty of murder in the first degree inferred a love interest in the case that was not directly evidenced. Only jurors favoring acquittal pointed to the fact that the defendant had a reputation for peacefulness. See also Pennington & Hastie: Evidence Evaluation, supra note 154, at 248-50.
African nations in the United Nations. Those subjects were first given a number determined by the spin of a modified roulette wheel (a number which none of them could have regarded as being related in any way to the true percentage). Subjects for whom the wheel yielded an initial number of 10 gave a median percentage estimate of 25%, while subjects whose spin landed on 65 gave a median guess of 45%. If people cling to initial positions determined by the spin of a roulette wheel, should it be surprising that it is hard to disabuse them of apparently reasonable causal theories?

Causal theories are apparently so congenial to human decision makers that there is substantial danger that once a causal hypothesis has been developed, no matter how tenuous its supporting evidence, decision makers will cling unreasonably to that theory. Given that this is the case, what is a wise decision maker to do: abandon causal reasoning or seek to improve that reasoning by subjecting it to more careful constraints? In my view this question answers itself based on the very evidence offered by the critics. I doubt that people can give up causal reasoning. If that is the case, then improvement rather than abandonment is the only realistic course.

J. Deficiencies of Pure Probabilistic Judgments

If probability, by itself, would give a better assurance of predictability, that might be a different matter, but in at least one important case, a pure probability model will lead to serious miscalculation. Even perfect co-variation, 100% probability, does

213. See Amos Tversky & Daniel Kahneman, in their introduction entitled Judgment Under Uncertainty: Heuristics and Biases, in HEURISTICS, supra note 27, at 14-15; HUMAN INFERENCE, supra note 27, at 41-42. See also Paul Slovic et al., Response Mode, Framing, and Information-Processing Effects in Risk Assessment, in DECISION MAKING, supra note 8, at 161-62.


215. One failure of probability to predict outcomes is of little interest here. Decision theory fails to give an accurate forecast when the improbable happens. Let me use a gambling analogy first. It is unlikely that a poker player will draw to an inside straight or that the craps player will “make the point” having failed to throw a seven or an eleven. See SALMON, supra note 141, at 195-96. Smart money would bet against both events. But in these forms of gambling, the smart money is wrong frequently enough to keep many players in the game.

For my purposes this phenomenon is relatively unimportant because, while probability forms a deficient base for prediction in such matters, causal reasoning as I have described it does no better. Isolating the “cause” of a particular roll of the dice would require a physics beyond that which will ever be readily available in the real world. So although those physics are clearly the cause, that cause is unknowable.
not guarantee causality. Let me give an example; a barometer moves up and down in fair correlation with changes in the weather. If the barometric pressure falls, assuming no change in elevation, bad weather is very likely. But no one would say that a falling barometer causes bad weather nor would one seek to avoid bad weather by pumping up the barometer.\textsuperscript{216} Rather people familiar with weather patterns recognize that an independent cause—the arrival of a low pressure system—produces both the fall in the barometer and the bad weather.\textsuperscript{217} This pattern of co-variation evidencing an as yet unsuspected outside cause goes by the name of the “conjunctive fork,”\textsuperscript{218} and correlations which ignore the underlying cause are labeled “spurious.”\textsuperscript{219} So even 100\% correlation, if it is spurious, can cause a decision maker to seriously misjudge the world through adopting the wrong causal argument. An example from the legal realm lies in the treatment of bail. Throughout most of American history, it was assumed that the threat of bail forfeiture caused criminal defendants to appear to face trial. Thus, bail was required in almost every case. A study done by the VERA Foundation\textsuperscript{220} in New York showed that pre-existing ties to the community, with or without bail, insured appearance.\textsuperscript{221} The ability to post bail, rather than causing

\textsuperscript{216} If the example of misguided attacks on barometers strikes you as far fetched, let me give you another example involving the same instrument. If all the decision maker knows is the correlation between bad weather and a falling barometer, she may conclude that every time she climbs a mountain with her barometer that rain is imminent. Because she lacks a causal theory about why barometers behave as they do, she does not realize that the loss of atmospheric pressure occasioned by the gain in altitude drops the barometer just as effectively as an approaching low pressure system. In fact, of course, barometers carried in airplanes are called altimeters and are prized precisely because of this property.

\textsuperscript{217} There is a class of instances of this type known as Newcomb's problems. See Ellery Eells, Rational Decision and Causality (1982); Robert Nozick, Newcomb's Problem and Two Principles of Choice, in Rationality, supra note 28, at 207; David Lewis, Causal Decision Theory, in Rationality, supra note 28, at 235.

\textsuperscript{218} See Salmon, supra note 141, at 158-68. Salmon credits Hans Reichenbach with coining the term “conjunctive fork.” Id. at 158. Problems of this type were used by Tversky and Kahneman to show people's predisposition to accept what may be a spurious causal explanation. See supra note 208 and accompanying text.

\textsuperscript{219} See Eells: Probabilistic, supra note 172, at 59 (defining a spurious correlation as one in which “neither [correlated] factor causes the other and the correlation disappears when a third variable is introduced and ‘held fixed’”).

\textsuperscript{220} The Foundation was created by industrialist Louis Schweitzer to address problems of pretrial detention. It was named for his mother.

\textsuperscript{221} See the report of Herbert J. Sturz, the executive director of the VERA Foundation, on the Manhattan Bail Project, in National Conference on Bail and Criminal Justice 43-51 (Robert M. Fogelson ed., 1974). As of the time of that report the “Project” had released 2630 criminal defendants without bond on personal recognizance (promise to return). Ninety-nine percent of those released voluntarily
return for trial, was itself a function of some forms of community ties. The entire bail system, as a vehicle for insuring presence, was based on a “spurious” causal judgment.

Such spurious correlations establish that probability is not an infallible guide to prediction. If decisions are governed by correlation alone, the decision maker may be misled into believing that the correlation will hold even when the true underlying cause is absent. A defendant with ready cash, but no community ties, may be likely to post bond and disappear.

One more thought, in passing, on this issue. Throughout this Article, I have been wrestling with the role of coincidence or the product rule for conjunctive probability for decision making. You will recall that in the decision theory model, the adoption of the product rule requires decision makers to treat coincidence as unlikely and strongly decrease the probability of the co-occurrence of two events. Cohen, in developing his theory of induction, proposed that a coincidence should be viewed as at least as likely as its least likely aspect. Coincidence did not drive down probability as it did in decision theory, but it did not materially contribute to understanding the phenomena. But, what is the conjunctive fork if not the epitome of coincidence? If two events, each rare by itself, always co-occur, the decision maker can conclude something affirmative based on that co-occurrence. It is highly likely that either one event causes the other or a third event causes them both. In any case, some causal relationship is almost certainly afoot. Rather than making us less willing to accept a co-occurrence as likely—as decision theory advises—coincidence may by its very rarity, when coupled with persistence, convince a decision maker that some powerful underlying explanation is at work.

The point will become clearer if I look at how the English language treats the word “coincidence.” In its simplest meaning the word signifies co-occurrence: the meaning of joint events attached by decision theory. But at a secondary level the word takes on either the naif or sardonic meaning of “a remarkable concurrence of events or circumstances without apparent causal connection.” The point is, of course, that the co-occurrence of events may in itself be sufficient evidence that a causal link

---

222. See supra Section II.E.
223. See supra Section III.B.
224. OXFORD, supra note 64, at 220.
exists between the phenomena involved. If such a causal link does exist, co-occurrence is more, rather than less, likely. The prime debate between decision theory and causal explanation lies in precisely this different treatment of some joint probabilities.

**K. Summary**

Causal reasoning has its risks as the cognitive psychologists cited here stress. Decision makers may be incorrect in perceiving correlations of supposed cause and effect. People have a tendency to see causal relationships when two phenomena are correlated and to disregard the issues of temporal precedence or the possibility of a third, unidentified causal element. Once a decision maker has adopted a causal explanation, she will be loath to relinquish it.

On the other hand, it seems very likely that decision makers will use causal reasoning despite these drawbacks. Also, in the case of the conjunctive fork, true causal analysis always provides a better guide than do probability judgments.

Therefore, it seems to me better to attempt to improve causal reasoning rather than to dismiss it as a flawed strategy. That improvement would include: greater care in co-variance analysis as exemplified by Mill’s tests; attention to the temporal relations of cause and effect; review of the supposed cause to insure that it fits the problem context; and establishing that the necessary link has occurred between the supposed cause and its result. Once viable possible causes are identified using those tests, a choice among those causal candidates should be premised on their coverage and simplicity. A cause so identified is unlikely to harbor the defects that the psychologists fear.

**V. Conclusion**

In a moment I will relate what is known of Clarence Darrow’s use of these methods of decision making, but let me first summarize the conclusions of this Article. Three methods of decision making commend themselves to lawyers seeking to predict uncertain events: decision theory, inductive probability, and causal reasoning. Each has its strengths and weaknesses.

Decision theory handles structured problems, like settlement decisions, well. For less structured enquiries like the decision as to whether to take a case, decision theory asks the practitioner to assume that she has identified all of the key variables and that she knows the degree of likelihood attached to each. The lawyer faced with such a decision may well feel uncomfortable with these demands. In that case,
the less structured, more tentative approach afforded by inductive probability may be more congenial.

Inductive probability gives up the controlled rationality of decision theory and is thus less satisfactory in areas in which problems are well-structured. But precisely because of its incremental approach to issues, it lends itself to questions where the relevant variables are not self-evident. In fact, the major undertaking of inductive reasoning is the attempt to insure, in the face of inherent uncertainty, that all relevant variables are examined. Inductive probability is thus both less dogmatic in its requirements for such certainty of total coverage and more constructive in its methods for securing such coverage. That feature is what makes it useful in less structured problems such as case acceptance.

Mathematically, inductive probability expressly rejects the assumption of decision theory that the probability of all competing events must sum to one. That is, it rejects decision theory's postulate that the decision maker knows all the contingencies and can assign probability values to them. This is the root difference between the two methods.

Both decision theory and inductive probability deal in likelihoods or probabilities. Causal reasoning goes beyond probability in an attempt to explain why events occur. It builds on a foundation of co-variance observations but adds a further analytical step. It is suited particularly to those problems in which the lawyer is trying to decide what motivates others, be they judges, jurors or parties. In this Article, I examined the decision to waive or to retain a jury in a criminal case. Causal reasoning enjoys frequent use because it may be the only way available to answer questions of this sort.

Causal reasoning departs from decision theory most notably in its treatment of the probabilities of conjoined events. The occurrence of two unlikely events may, in that system, have a higher probability than the solitary occurrence of one alone. The prime debate between decision theory and causal explanation lies precisely in this different treatment of some joint probabilities.

Which then, if any of these methods, did Darrow use in deciding his strategy in the Leopold and Loeb case? There is no source of which I am aware in which the lawyer privately, and therefore candidly, set forth his thinking. But Darrow did address the issue in his closing argument itself. Baited by the prosecution for his unwillingness to trust the case to a jury, Darrow acknowledged that he had reasons for preferring a bench hearing. It is unlikely that, in the course of carefully crafted advocacy, Darrow shared his entire decision process with his adversaries and the Court, but this is what he said:

We did plead guilty before Your Honor because we were afraid to submit our case to a jury. I would not for a moment
deny to this court or to this community a realization of the serious danger we were in and how perplexed we were before we took this more unusual step.

I can tell Your Honor why.

I have found that years and experience with life tempers one’s emotions and makes him more understanding of his fellow-man.

... ...

I am aware that as one grows older he is less critical. He is not so sure. He is inclined to make some allowance for his fellow-man. I am aware that a court has more experience, more judgment and more kindliness than a jury.

... ...

I know perfectly well that where responsibility is divided by twelve, it is easy to say: “Away with him.”

But, Your Honor, if these boys hang, you must do it. There can be no division of responsibility here. You can never explain that the rest overpowered you. It must be by your deliberate, cool, premeditated act, with no chance to shift responsibility.

... ...

Now, let us see, Your Honor, what we had to sustain us [in making the decision to plead]. Of course, I have known Your Honor for a good many years. Not intimately. I could not say that I could even guess from my experience what Your Honor might do, but I did know something. I knew, Your Honor, that ninety unfortunate human beings had been hanged by the neck until dead in the city of Chicago in our history. Some ninety human beings have been hanged in the history of Chicago, and of those only four have been hanged on the plea of guilty—one out of twenty-two.

I know that in the last ten years four hundred and fifty people have been indicted for murder in the city of Chicago and have pleaded guilty. Four hundred and fifty have pleaded guilty
Lawyer Decision Making

in the city of Chicago, and only one has been hanged! And my friend who is prosecuting this case deserves the honor of that hanging when he was on the bench. But his victim was forty years old.

Your Honor will never thank me for unloading this responsibility upon you, but you know that I would have been untrue to my clients if I had not concluded to take this chance before a court, instead of submitting it to a poisoned jury in the city of Chicago. I did it knowing that it would be an unheard-of thing for any court, no matter who, to sentence these boys to death.

And, so far as that goes, Mr Savage [the prosecutor] is right. I hope, Your Honor, that I have made no mistake.225

From this speech, it seems to me that Darrow used at least two of the three decision methods discussed here and may have used them all. He does not discuss a decision theory calculation, but it would be very surprising in this context if he laid out for the Court his decision tree. Yet it is clear that he has done some probability calculations for he knows that of people hanged in Chicago throughout its history, less than 5% had plead guilty. And he knows that, within the last ten years, only 2% of those pleading guilty were hanged. These likelihood data could figure in either decision theory or inductive probability calculations.

It is also clear that Darrow is engaged in causal reasoning. Age promotes mercy he says, and it is more difficult to kill when the responsibility rests with a single decision maker than when it is distributed among twelve.

While I have no certainty as to how lawyers in general go about decision making, I could surely do worse than draw to your attention methods which worked for such an able advocate in such treacherous circumstances.

225. DAMNED, supra note 2, at 23-25.